



Leporini, B., Buzzi, M. and Hersh, M. (2022) Video conferencing tools: comparative study of the experiences of screen reader users and the development of more inclusive design guidelines. *ACM Transactions on Accessible Computing*, (doi: 10.1145/3573012).

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

© Association for Computing Machinery 2022. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *ACM Transactions on Accessible Computing*. <https://doi.org/10.1145/3573012>.

<https://eprints.gla.ac.uk/287267/>

Deposited on: 22 December 2022

Enlighten – Research publications by members of the University of Glasgow_
<http://eprints.gla.ac.uk>

Video Conferencing Tools: Comparative Study of the Experiences of Screen Reader Users and the development of more Inclusive Design Guidelines

Barbara Leporini *

ISTI-CNR, via Moruzzi 1-I56124 Pisa, Italy, barbara.leporini@isti.cnr.it

Marina Buzzi

IIT-CNR, via Moruzzi 1-I56124 Pisa, Italy, marina.buzzi@iit.cnr.it

Marion Hersh

University of Glasgow, Glasgow G12 8LT, Scotland, marion.hersh@glasgow.ac.uk

Since the first lockdown in 2020, video conferencing tools have become increasingly important for employment, education, and social interaction, making them essential tools in everyday life. This study investigates the accessibility and usability of the desktop and mobile versions of three popular video conferencing tools, Zoom, Google Meet and MS Teams, for visually impaired people interacting via screen readers and keyboard or gestures. This involved two inspection evaluations to test the most important features of the desktop and mobile device versions and two surveys of visually impaired users to obtain information about the accessibility of the selected video conferencing tools. 65 and 94 people answered the surveys for desktop and mobile platforms respectively. The results showed that Zoom was preferred to Google Meet and MS Teams, but that none of the tools was fully accessible via screen reader and keyboard or gestures. Finally, the results of this empirical study were used to develop a set of guidelines for designers of video conferencing tools and assistive technology.

CCS CONCEPTS • Human-centred computing~Accessibility~Accessibility design and evaluation methods • Human-centred computing~Accessibility~Empirical studies in accessibility

Additional Keywords and Phrases: Videoconferencing tools; blind people; screen reader; screen reader users; guidelines; evaluation methodology; accessibility; survey

ACM Reference Format:

Barbara Leporini, Marina Buzzi and Marion Hersh. 2021. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In

*** Place the footnote text for the author (if applicable) here.

1 INTRODUCTION

Video conferencing systems have been in use for a number of years [55], but it is only in the last two years that they have become part of everyday life for many people. From the start of the first lockdown, video conferencing systems have allowed us to carry out numerous activities from home, such as distance learning in schools and universities, office and many other types of work, and social interaction [33], [44], and [59] that otherwise would not have been possible due to restrictions on face to face activities. These tools have evolved over the last two years, including the development of additional features and interface functionalities. However disabled people, particularly those who use assistive technologies to interact with computers and mobile devices or have other accessibility requirements, may experience difficulties accessing and using them [18], [30], [34].

Computers and technology in general are often the key enabling factors for blind people to work and achieve social integration [10]. They therefore need to be fully accessible and usable for all. Despite the existence of numerous accessibility guidelines to help developers and designers create more accessible applications and web pages [61] and [62], disabled users, including those who interact via screen reader [57], are still likely to experience barriers to their use [52]. Studies show that the guidelines do not cover all the problems encountered by disabled users [16], [51]. Visually impaired screen reader users regularly experience numerous difficulties when interacting with graphical interfaces [28]. Mobile devices may introduce additional accessibility issues when the touchscreen is accessed via screen reader [15], [22]. It is therefore important to investigate how screen reader users use video conferencing tools. Popular video conferencing systems claim to meet the main W3C accessibility guidelines with only a few exceptions [39]. While meeting accessibility guidelines should be strongly encouraged, studies have shown that this is not necessarily sufficient to ensure good usability by screen reader users [37], [40], [49], [60]. Consequently, end-user testing is also required. This work aims to make a contribution in this context.

The study was motivated by an interest in improving blind and partially sighted people's wellbeing and ability to participate on equal terms and have equal opportunities. The importance of video conferencing tools since the first lockdown and their continuing use make their accessibility and usability an important component in ensuring the well being and participant of blind people. This led to two separate research questions relating to the ability to use (accessibility) and ease of use/satisfaction (usability). The questions are as follows:

- 1) To what extent are screen reader users able to use all the features of videoconference tools or do they experience barriers to accessing some important functions via screen reader?
- 2) How easy and satisfying is it for screen reader users to use videoconference tools? In particular, is interaction easy or do users experience cognitive overload and a poor and time-consuming interaction?

We answered these questions in this study by investigating screen readers' experiences when interacting with three commonly used video conference tools for desktop and mobile platforms: Zoom, MS Teams and Google Meets. The approach included an inspection evaluation and analysis by accessibility experts on tool use via screen reader and keyboard (desktop) or gestures (mobile). This was followed by surveys of the experiences of blind users of these tools on desktop and mobile devices and any resulting accessibility and usability issues to investigate and corroborate or modify the results of the inspection analysis This work is an extension of the study presented in [39]. It includes responses from additional participants, responses to versions of the survey for both mobile and desktop devices and a set of guidelines for designers. More specifically:

- (1) The 29 responses reported in [39] for the three tools on desktop/online platforms have now been increased to 65.
- (2) The study of desktop/online versions of three tools in [39] has been supplemented by a study of the same three tools on mobile platforms (i.e. via touchscreen devices) using the same methodology. 94 people responded to the survey on mobile video conferencing tools.
- (3) A set of guidelines for designers is proposed which draws on the accessibility and usability issues identified in the two studies for screen reader users of video conferencing tools using a keyboard (desktop) or gestures (mobile).

The paper is organized into 9 sections. Section 2 briefly discusses the relevant literature; section 3 introduces the context; and section 4 presents the methodology. Sections 5 and 6 respectively present the results of the inspection evaluation and the surveys. The results are discussed in section 7 and a set of guidelines for designers, aiming at improving the usability of videoconferencing tools, is presented in section 8. The paper ends with conclusions in section 9.

2 RELATED WORK

Due to the COVID-19 pandemic, universities and schools around the world were physically closed and distance learning became a daily reality. This move from face-to-face to digital education forced both teachers and students to adapt to a new reality, with associated difficulties [20]. Schools worldwide quickly switched to online learning adopting a variety of virtual learning environments (VLEs), such as Google Classroom, Moodle, and Blackboard to deliver virtual lessons and content [54] and/or videoconferencing tools such as Zoom [56]. In educational environments learning management systems have developed from static tools to include live virtual classrooms which enable remote interactions [32] and cloud-based learning environments have been increasingly popular with pedagogical benefits [21], [43]. Tools such as Zoom enable students to interact with teachers and classmates via smartphones, tablets and/or computers. Best practices for delivering lessons and supporting learning in virtual classrooms have been discussed in numerous studies [7], [8], and [54]. It is crucial that educational content and VLEs are fully accessible to disabled students and staff. A review analyzing 14 studies in educational environments revealed many problems [54]. They included reductions in academic performance due to the lack of a quiet place to study, distractions, lack of adequate access to course materials, inadequate digital skills (of both students and teachers), and insufficient network bandwidth for a good Internet connection [7], [20], [54]. However, the focus of all these papers was mainly the accessibility of the educational and pedagogical content rather than the VLE or videoconferencing tools.

Zoom is one of the most commonly used videoconferencing tools worldwide, particularly in the USA. An online study of 31 university students' experience with Zoom indicated that they were not fully satisfied with their learning experience during this pandemic period. They considered that the main advantages were the flexibility of attending classes anywhere, easier interaction, written communication, and the use of multimedia. The main problems were distractions (which were common at home), poor quality interaction and feedback, poor education quality and technical difficulties [56]. However, these studies do not explicitly mention and may not have included disabled students who may experience even greater difficulties than their non-disabled peers.

When interacting with videoconference systems, disabled participants can experience a variety of accessibility problems and challenges. A combination of automatic web accessibility evaluation tools and manual evaluation have been used to assess the accessibility features of three video conferencing tools in different interaction scenarios in an educational context for disabled students. Zoom was found to be the most accessible of the three tools followed by Big Blue Button and then Webex [23].

Two recent studies have investigated the accessibility challenges of video conferencing technology for the Deaf community. One study analyzed many aspects of the interaction of deaf people with the video conferencing system Zoom, highlighted barriers and challenges and suggested accessibility improvements for this population [4]. The other study used interviews and co-design sessions to investigate the main accessibility barriers for Deaf and hard of hearing people who used a national sign language to communicate [53]. They found that popular video conferencing platforms do not adequately consider communication and collaboration for Deaf signers and developed guidelines for conducting online research with them.

The requirements of blind users who do not have access to visual content and navigate via keyboard were first considered in [39]. Accessibility considerations should cover tool use by both participants and hosts, and the accessibility of user interfaces and shared content. Blind people can miss important information and communications when this is provided in purely visual form. When analyzing meeting accessibility for blind people, it is important to note that meetings include visual content and nonverbal communication elements such as gestures, facial expressions and dynamic changes of object focus [49]. The authors investigated making visual tools for brainstorming accessible and the detection and delivery of non-verbal communication cues. This is important for both face-to-face and online meetings.

Research on the accessibility of video conferencing tools is still in its infancy. Videoconferencing platforms include a set of tools that are intended to function together seamlessly but do not necessarily do so when used by blind, other disabled or older people. They offer a wide range of functions, including muting/unmuting the microphone, turning the video on/off, hand raising, sharing the screen or files and playing videos. However, tool design has not necessarily fully considered the diversity of users and modes of access. Thus, blind people may experience difficulties with 1) exploration and navigation, i.e. moving rapidly between different areas and tools, and 2) proficient and speedy use of functions such as hosting a meeting, adding participants, accepting a meeting invitation and controlling devices such as speakers and video-cameras. The available tools for navigation and exploration include access keys, tab keys, and screen reader commands.

The Worldwide Web Consortium (W3C) has published a draft document on the Accessibility of Remote Meetings [63]. In addition to (standalone client) conferencing systems, the W3C group analysed several specialized remote meeting systems, including Conference/event, Educational (Learning Management Systems), Medical (eHealth) and XR (immersive remote meeting) platforms [63]. However, videoconferencing tools are often preferred by disabled participants, especially blind ones, as they are better integrated with the assistive technology they are using (as discussed later in this paper).

There is legislation on making information and communication technologies (ICT) accessible to disabled people. This includes, in Europe, Directive (EU) 2016/2102 requiring public sector bodies to make their web sites and apps accessible to disabled people and the European Accessibility Act (Directive 2019/882) aimed at fueling the EU market for accessible products and services by overcoming the different rules of Member States. In Europe the standard EN 301 549 for digital accessibility has been defined by ETSI, the European Telecommunications Standards Institute, (<https://www.etsi.org/>).

In US the Rehabilitation Act requires federal agencies to make their ICT accessible to disabled people. US legislation encourages the use of voluntary product accessibility templates (VPATs) stating the list of accessibility requirements to support developers in ensuring compliance with them. Zoom and Google Meet have publicly available VPATs and Microsoft provides detailed accessibility information on Teams (access key, configuration options, features, etc.) although the VPATs were not found online.

The Google Meet VPAT shows full support for most of the access criteria. However, two criteria related to screen reader interaction are only partially supported, namely 2.1.1. Keyboard and 2.4.3 focus order. In addition, there is only a web version of Google Meet, whereas Zoom and Teams can also be downloaded and locally installed as plug-ins or native desktop apps. These versions are generally more accessible via screen reader than the web versions. In addition, the provision of more options generally improves accessibility.

Zoom considers itself to be compliant, with some exceptions, with standards including the W3C Web Content Accessibility Guidelines (WCAG) 2.1 AA, the revised Section 508 standards and the EN 301 549 standard. It provides downloadable VPATs for the different versions: Zoom Client App for Windows and macOS, Zoom Mobile App (for iOS and Android), Zoom Rooms and Controller, Zoom Plug-ins for Outlook for Mac OS and Windows, Web Pages and Extensions).

While this is a positive development, it is not necessarily sufficient to provide full accessibility for all disabled people. In addition, the vendors' product web pages indicate that none of these tools is defined as fully accessible via screen reader, as there are a number of exceptions for several functions [39]. The small number of existing studies confirms that videoconferencing systems are not fully accessible. Even popular chat systems such as WeChat, Hangouts, Tango, Line and Viber are not fully accessible to blind people [42]. A preliminary study on the accessibility of several popular video conferencing tools for disabled people was carried out by [29]. However, this is a preliminary rather than a full study, and a more systematic study of the experiences, needs and suggestions of blind people is required. Other studies have expanded knowledge in this field, such as [24], [39]. However, specific guidelines based on user experience have not yet been proposed.

The W3C WAI has produced draft accessibility guidelines for developers and people choosing software for remote meetings [63]. This document reports and summarizes the W3C accessibility guidelines relevant to the selection and development of remote meeting software tools and platforms to support users' access needs. Section 3.2 Creating accessible remote meeting software platforms requires software developers to ensure accessibility features and support for accessible user interfaces when designing and maintaining remote meeting systems. W3C provides several accessibility resources as well as other guidelines. This is an important initiative, but the guidelines are broad based and, for instance, lack sufficient details of accessibility and usability for screen reader users.

3 BACKGROUND

3.1 Video Conferencing Tools

Video conferencing tools enable multiple people, who are generally in different locations, to meet and collaborate at a distance without travelling by sharing audio, video, text and presentations in real time over the internet or other networks. Due to the Covid-19 pandemic, the use of video conferencing tools has significantly and possibly exponentially increased in both employment and education [17]. Some remote workers have experienced increased productivity and an enhanced work-life balance [9], [13]. Others have missed face to face interactions with colleagues, found virtual meetings more tiring than face to face ones and had a poorer work-life balance as it is more difficult to make the distinction between work and other activities when they take place in the same location [31]. In some cases, users with particular accessibility requirements may use audio only tools e.g. phone call to access the meeting.

Video conferencing systems have become increasingly enriched with functions and features to deliver a better experience. However, this may reduce rather than improve accessibility for disabled users, including for basic functions. There are two main roles for managing different tasks and functions:

- a) Host: hosts create and manage the online meeting. They include teachers and tutors who set up and facilitate online learning activities with groups of students, including lectures, classes and group work on particular activities, and the organiser or convenor of a group of participants when the tool is used in meetings and conferences. Hosts also include students who organise online learning (or social) activities for themselves and their teachers as part of active learning.
- b) Participant: participants engage in meetings set up and chaired by others. They include students taking part in online learning activities and participants and collaborators taking part in a wide range of meetings and conferences.

The focus of this work is the “participant” role, as it is considerably more common.

3.1 Screen Reader Interaction

A screen reader (SR) is an assistive technology used by visually impaired people to interact with a computer or mobile device, such as a smartphone. The screen reader mediates between the user and the operating system (including its applications), assisting individuals by interpreting the user interface which is read aloud sequentially by means of a voice synthesizer, or written on a Braille display [38]. Therefore, screen readers have become the most appreciated tool for blind people, despite requiring a certain amount of effort by the user to learn how to use them proficiently, including advanced commands.

Several screen readers have been developed over the last few years, including: Jaws for Windows (<http://www.freedomscientific.com/>), NVDA (<https://www.nvaccess.org/>), for Windows, and VoiceOver (https://www.apple.com/voiceover/info/guide/_1121.html) for IOS and OS X, and Talkback (<https://support.google.com/accessibility/android/answer/6283677>), for Android-based devices.

The screen reader announces the content displayed on the user interface once it appears on the screen, or on-demand, i.e. when the user wants to read a specific UI portion or component through a given command or by moving the system focus. For example, when a dialog box is shown on the screen, the screen reader usually reads its content automatically. Similar behaviour occurs when a web page is loaded. Also, when the user navigates via keyboard through the Tab or Arrows keys, the system focus moves on to the interactive components (i.e. the elements with which the user can interact) such as buttons, textboxes or combo-boxes, radio buttons or checkboxes, or links, and the SR announces the object the focus is on. For instance, while navigating the toolbar by using the Alt and Arrows keys, each item is immediately read by the SR. When instead the user wants to get the title of the current window, or read the status bar, specific SR commands have to be used (e.g. the Insert+ T command to read the window title via Jaws and NVDA screen readers). The interaction via screen reader on mobile platforms is very similar to the desktop one, except that gesture-based commands are used instead of keyboard commands.

To sum up, what is read automatically by the SR, or activated by the user with specific SR commands or by navigating the user interface, depends on interface features and the user context. The SR behaviour can vary when interacting with a web page rather than a desktop or a mobile application. The result of the keystrokes/gestures varies depending on the platform, the type of screen reader and the application it is interacting with ([5], [6], [25], [27] and [46]). Understanding how SR work and how screen reader users interact with user interfaces (UIs) can aid designers and developers in carrying out accessibility inspection and evaluation.

4 METHOD

Our methodology combined two well accepted approaches: (1) expert evaluation and (2) end-user surveys. This had the advantage of considering both the end-user and expert perspectives. Expert evaluation can be used to help designers by identifying areas of the design that need particular attention. In particular, it can contribute to (a) highlighting tasks that users are likely to have difficulties in completing, and (b) identifying interface components that may require attention to assure a good interaction. We also recognise the importance of the direct experiences of end-users, in this case screen reader users, as they are the only people who know what is and is not accessible to them. We consider it unlikely that the use of self-reported results will introduce bias since participants are unlikely to have reasons to report incorrectly and even more unlikely for large numbers of them to do so in a particular direction. The end-user survey enabled us to further investigate the accessibility and other issues identified by the inspection evaluation through consideration of the experiences of screen reader users.

Most videoconferencing tools offer both desktop and mobile versions, which generally use different operating systems. Therefore the study considered both these versions through two separate end-user surveys, one of desktop and the other of mobile devices, and different approaches to the inspection evaluation of the desktop and mobile platforms to take account of the differences in screen reader interaction mentioned in section 3.1.

Analysis of the study results and, in particular, the problems experienced and examples of good practice were used to develop guidelines for the developers of video conferencing tools.

4.1 Expert Evaluation

There are three main approaches to expert accessibility evaluation: (1) using semi-automatic tools ([1], [3], [48]); (2) heuristics-based evaluation involving experts ([26] and [47]); and (3) end-user testing by (screen reader) users ([12], [19] and [58]). However, the focus has generally been on testing heuristics by checking whether particular criteria were implemented. For instance, [2] checked the compliance of six video conferencing tools with WCAG 2.1 and 2.2 accessibility guidelines.

Our methodology goes beyond this approach in aiming to carry out an expert evaluation of screen reader user behaviour. It combines expert evaluation with testing by screen reader users. We have avoided the use of semi-automatic tools because (1) these are more oriented toward web pages/applications and (2) semi-automatic tools may not be able to detect important features such as ease of keyboard interaction or the availability of gestures, the number of steps to be performed to find and activate a function, and how the screen reader interacts with and detects the user interfaces. In fact, we do not want to limit our evaluation to detecting whether or not a button has alternative text or whether a function is simply keyboard-activated. Instead, we want to assess how the user manages to interact with the interface in a more practical and effective way. For instance, we want to consider the number of steps to be performed to find or activate a function. It is not always appropriate to use a shortcut to access a function to make it easy for the user to use and therefore accessible from the keyboard. Learning many shortcuts can be a considerable cognitive load for the user. Not all users are willing to learn numerous key combinations for many applications. It is therefore important to consider whether the interface can be navigated from the keyboard with only minimal effort by the user.

The inspection evaluation was carried out by two of the three authors, all of whom have a good understanding of the accessibility and usability of user interfaces for blind people. We consider expert evaluation by blind screen readers users who are also experts the ideal approach. Unfortunately, there are few such experts available. One of our two expert evaluators has been blind since childhood and is an expert screen reader user. The study involved the mobile versions of

the three tools, the desktop versions of Zoom and Teams and the web-based version of Google Meet, as it does not have a desktop version.

It involved a cognitive walkthrough [41] to detect the main issues encountered by screen reader users when carrying out tasks and critical aspects of these tasks. The screen reader Jaws for Windows 2021.2012.48 was used to interact with the user interface in the desktop environment and an iPhone X smartphone with IOS 15.5 and VoiceOver screen reader, with the Safari browser in the mobile environment.

In this study the inspection evaluation focused on identifying the crucial tasks in which a screen reader user may encounter difficulties when interacting with video conferencing tools. Only the basic functionalities of the ‘participant’ role were considered.

Using a video conferencing tool effectively requires participants to be able to do the following:

- (1) use the input devices (turn them on/off);
- (2) access status information e.g. check which devices are on/off and obtain information about the other participants;
- (3) participate actively in the meeting e.g. comment verbally, share screen content, and read/write chat messages. The user should be able to obtain a variety of data and access the tool status. This is relatively easy for non-disabled users who are able to obtain information visually from the interface, but may be more difficult for screen reader users who do not have a general overview of the interface and interact via the keyboard.

The functions (and tasks) considered in the evaluation are listed in table 1. They have been classified as ‘action’ tasks i.e. actions related to active tasks such as turning the microphone (mic) or video camera (cam) on/off; and ‘awareness’ tasks which mainly involve status checking activities, such as checking the on/off status of the microphone or video camera.

Table 1: Tasks considered in the evaluation

Type	Description
Action	F1. Joining a meeting
Action	F2. Hand raising (asking to speak in a meeting)
Action	F3. Turning the microphone on/off
Action	F4. Turning the video camera on/off
Action	F5. Audio, video, file and screen sharing
Action	F6. Using the chat
Awareness	F7. Accessing the shared content
Awareness	F8. Checking the microphone and video camera status
Awareness	F9. Obtaining participant information (number, names, who is joining or leaving)

4.1.1 Desktop Inspection Methodology

Desktop applications and web pages require screen reader (SR) interaction, generally via keyboard and special SR commands (see 3.1). In particular, the screen reader follows the system focus which can be managed via keyboard. Thus, the application needs to offer good system focus management and the inspection evaluation should include focus management. More specifically, evaluating each task via screen reader and keyboard involves consideration of the following:

- Use of the Tab and Arrow keys to move the focus onto the function and to explore the user interface and enable us to analyze screen reader behaviour when handling the focus. This interaction mode is very important, particularly for novice users who usually navigate through basic keys, such as the Tab and Arrows keys.
- Use of the user interface shortcuts to activate functions. This simplifies screen reader and keyboard interaction and saves time and effort by reducing a potentially large number of keyboard commands to a single one. This is particularly important for tasks that are carried out frequently.
- Use of screen reader commands for specific actions to get information on the user interface and function feedback (which non-disabled users generally obtain from the screen). This includes reading the window title and the status bar. These commands can be used when the focus does not work, there are no shortcuts or they do not work.
- Analysis of the clarity and relevance of the screen reader feedback messages, both when using the Tab key to move the focus and when using shortcut keys. This is particularly useful for blind people unable to see the user interface and its status when performing and triggering certain actions. Clear, accurate feedback about the current status and task success/failure is very useful, whereas irrelevant or ambiguous messages can lead to confusion.

Particular attention was given to screen reader feedback messages to users, to evaluate whether blind screen reader users are able to perceive what is happening and obtain visually displayed information via a screen reader (or Braille display) [11]. For instance, feedback that 'the microphone is on' or 'the microphone is unmuted' is easy to understand and does not require interpretation. 'Turn the microphone on' requires interpretation and more cognitive effort, particularly in meetings, as it requires interpretation. 'Turn on' or 'activate' is ambiguous, as it is unclear whether it refers to the microphone or camera.

4.1.2 Inspection Methodology of Mobile Devices

Blind users generally use gestures on touchscreen devices, as there is rarely a hardware keyboard available. Mobile interfaces are also simpler than desktop ones due to the smaller screen. In fact, the layout of virtual classrooms on mobile devices (and mobile apps in general) requires reduced user interface information due to the smaller screen [33].

The methodology needs to consider the following:

- Use of the Left and Right flick gestures to move the focus onto the function and explore the user interface and enable us to analyze screen reader behaviour when handling the focus;
- Use of the double Tap to click onto an interface element, i.e. on an interactive component (e.g. buttons, checkboxes, menu, etc.);
- Detection of the positions of the main buttons to help determine whether functions such as turning on/off camera and microphone are within easy reach;
- Analysis of the clarity and relevance of screen reader feedback messages, both when using the right and left flick gestures to move the focus and when using a double tap to click on an item. This includes icon and button labels, and any other text content shown in the interface. As with desktop systems, clear, accurate feedback on the current status and task success/failure is very useful in helping SR users to understand what is happening, whereas irrelevant or ambiguous messages can lead to confusion.

4.2 End-User Surveys

A mixture of quantitative and qualitative data was collected using two questionnaires, both divided into three sections:

- I. user demographic data: five questions

- II. tool use: a separate subsection for each tool to make it easier to focus on the particular tool and allow participants to skip sections for tools they have not used. The desktop survey had four questions in each subsection and the mobile device survey had six. These questions are presented in an appendix. Both questionnaires asked about the ability to access a list of functions and how satisfied users were with the tool. The desktop questionnaire also asked how often shortcut keys were used and how easy it was to use the tool without them. The mobile device questionnaire also asked how participants accessed the tool features; how useful they found the automatic reading of screen content and their preferences for desktop, mobile and web versions of video conferencing tools.
- III. comments and suggestions: a single question asking for comments, descriptions of issues and suggestions for improving the three tools.

Simple language was used and questions were formulated to make them easy to remember to facilitate questionnaire completion with a screen reader. This allowed users to use key presses (e.g., “h” for the Jaws screen reader) to move to the next or previous question and immediately listen to the question number. The question title was kept short, to avoid annoying the user by repeatedly reading long heading elements. This navigation structure allowed users to get to each question quickly and read only its number and a very short description. They could then read the whole question by using the arrow keys and move directly to the next question by pressing the command key. All the questions on the tools were optional. This reduced the stress of users having to find unanswered questions, but could risk some relevant questions not being answered.

Each questionnaire was made available on web-based forms using Google Docs. They were piloted with two blind people before being distributed throughout the visually impaired community in Italy using email and general and specific mailing lists of the Italian Association for the Blind.

Both quantitative and qualitative data were obtained from the end-user survey. Statistical analysis of the quantitative data included calculation of percentages, averages and standard deviations and the use of χ^2 to calculate statistical significance. Analysis of the qualitative data involved one stage of coding to identify the main themes, as well as searching for potential explanations of outcomes of the statistical analysis. It should be noted that further research will be required to investigate these explanations and confirm, modify or disprove them.

5 RESULTS OF EXPERT EVALUATION

5.1 Desktop Inspection Results

Table 2 summarizes the results of the inspection evaluation for Zoom, Google Meet and Teams. For each function ‘yes’ in the relevant cell is used to indicate (1) the focus is supported, (2) there is a shortcut for the function, (3) screen reader (SR) feedback is appropriate and a blank that the function is not available, inaccessible or not supported. The term “partial” is used to indicate a) the process to reach an object requires many steps or is complex b) performing a task or function is not intuitive or is too complex.

Table 2: Inspection evaluation of desktop Zoom, Meet and Teams: accessibility features

	Focus			Shortcut			SR feedback		
	<i>Zoom</i>	<i>Meet</i>	<i>Teams</i>	<i>Zoom</i>	<i>Meet</i>	<i>Teams</i>	<i>Zoom</i>	<i>Meet</i>	<i>Teams</i>
F1. Joining a meeting	Yes	Yes	Partial				Yes	Yes	Partial

F2. Raising hand	Yes	Yes	Yes	Yes			Yes	Yes	Yes
F3. Turning mic. on/off	Yes	Yes	Yes	Yes	yes	Yes	Yes	Yes	Yes
F4. Turning cam on/off	Yes	Yes	Yes	Yes	yes	Yes	Yes	Yes	Yes
F5. Screen sharing	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
F6. Using the chat	Partial	Partial	Yes	Yes			Yes	Partial	Yes
F7. Accessing shared content				Partial					Partial
F8. Checking mic&cam status	Yes	Yes	Partial				Partial	Yes	Partial
F9. Obtaining participant info	Yes	Yes	Partial	Yes			Yes	Yes	Partial

F1. Joining a Meeting.

This action can be carried out relatively easily via keyboard and screen reader for Zoom and Google Meet. However, the procedure on Teams is quite complicated. The focus cannot be moved onto the button “participate” displayed on the screen, and the SR feedback is not useful. Joining a meeting requires multiple key presses: moving the focus via the Tab key onto the chat list (requiring multiple Tab key presses), using the arrow keys to select the message announcing that “the meeting is starting on...”, pressing the Enter key twice to detect the “participate” button, and finally joining the meeting by pressing the Enter key yet again. This is hardly quick and intuitive. Another option is using SR commands to activate a virtual cursor to enable the user to explore the user interface as a web page. This allows the user to find the “participate” button and navigate many other commands. However, the user interface is quite confusing and novice and non-expert users are unlikely to be aware of this option which is by no means intuitive, or particularly easy to use.

F2. Asking to speak in a meeting.

This function is accessible in Zoom and Teams, but not in Google Meet. At the time of the surveys, it was only available for the G Suite business and education versions of Google Meet, but has subsequently been released for all versions. Only 10 users (36%) were able to use this function and seven commented on the need for additional shortcuts. Thus, screen reader users will need to make a verbal request to speak in Google Meet. They may not want to draw attention to themselves in this way and it can be difficult to find a suitable break in the meeting speech to do so. Only Zoom provides a shortcut for hand raising or lowering. The SR feedback is clear for all three tools.

F3 and F4. Turning the microphone and video camera on and off

These functions are fully accessible for the three tools tested. The focus can be moved onto the specific button, or can be used with the assigned shortcut. The SR feedback is satisfactory in Teams and Google Meet, but Zoom feedback that an action has been carried out requires a specific screen reader script to be loaded.

F5. Audio, video, file and screen sharing.

This function is used for video conference presentations and in meetings to give participants access to information. The screen reader or keyboard can be used for the “share” function with either the entire screen or a specific window selected. Sharing the screen or a window and including the audio (when supported) was straightforward, but there were issues with file sharing. In Teams this required the File Tab, which could not be easily detected with a screen reader. In

Meet only the host but not participants had access to this function and could attach files to be shared in the meeting. Zoom used the Tab key and File sharing could be detected by the screen reader when exploring the chat area.

F6. Accessing shared content.

F5 sharing allows users to share content with other participants, whereas F6 accessing shared content allows users to access content shared by other participants. Unfortunately, all three tools show screen content shared by other users as an image, making it inaccessible for screen reader users. SR feedback is just the message “screen sharing by the speaker”. However, screen readers are able to detect shared PowerPoint files in Teams. In this case, the PowerPoint file content can be detected by a screen reader if the presentation has been designed to be accessible, including titles for all slides and alternative descriptions for all images. The shared content can be accessed by using the Tab key to move the focus.

F7. Using the chat.

To use the chat users need to be able to write messages and read and edit other people's messages. Zoom and Teams tools provide better support for using the chat than Google Meet. Only Zoom provides a shortcut to move into the chat area. In Teams the chat area can be reached with the Tab key and then opened using the “Show the chat” button. It can also be accessed in Teams via the focus. The user can then use the Tab or Arrow keys to move between reading messages and the edit box for writing them. This allows screen reader users to read and write text messages. However, a shortcut to open the chat area would be very useful and has been requested by several users, or instance “It would be useful to have keys to get to ... the chat ... faster’. Keyboard access to the “Chat” button and edit box in Google Meet is possible, but difficult with a screen reader. The lack of a shortcut to open the chat area means that a large number of steps are required and it can be difficult to carry them out and listen to the meeting at the same time. Reading messages is also more difficult in Google Meet. The screen reader is able to detect some messages, but the list of messages can only sometimes be detected, almost as if it is appearing and disappearing.

All the tools enabled users to read aloud a message written by another participant when the tool window has the focus. This is very useful, as the message is automatically read by the screen reader. However, users may find it difficult to listen to messages and speakers at the same time and therefore a function to scroll through or search previous messages would be very useful. It is very difficult to use links or copy a message in the chat in all the tools. This means the chat is considerably less useful as a source of information than it could be.

F8. Awareness of microphone and video camera status.

Users often need to know if their microphone and video camera are on or off. Unfortunately, this information cannot be obtained directly by screen reader users and has to be inferred. For instance, the user can turn the microphone or the video camera on/off by using the shortcut and obtain its status from the SR feedback. However, this is not a direct approach and requires the device status to be changed and possibly changed back again to the desired status. Alternatively, the user can move the focus to the on/off button and obtain the current label of this button. A “turn on” label means the device (microphone or video camera) is currently turned off and vice versa. However, this requires interpretation and could require numerous steps to be performed, possibly distracting the user from focusing on the speakers and the meeting.

F9. Information about participants (number, names, who is joining or leaving).

Screen reader users should have access to the same information about the other participants, including their names and the number of them, as non-disabled participants. Zoom, but not Teams and Meet, provides a shortcut to open the participant list. All three tools provide access using the Tab key (Zoom and Teams) or the Arrows key (Meet), but this approach generally requires several steps. The Arrow key can then be used to access information about a specific participant. Zoom provides the most complete and clearest information with all data presented in a meaningful order.

Teams provides similar information for each participant, but the information provided may require interpretation or the order may be inappropriate. Meet provides similar information, but it requires more interpretation and is more difficult to read. Table 3 shows an example of the different ways in which participant information is read by the three tools.

Table 3: Participant information read by the screen reader in the three tools

Tool	Participant information
Zoom	Alex Smith, computer audio unmuted, video on, hand raised Bob White, computer audio muted, video off
Google Meet	Alex Smith Turn off Alex Smith's microphone <i>button</i> Bob White Turn off Bob White's microphone
Teams	Alex Smith's profile picture, Alex Smith, on the phone, hand raised, unmuted Bob White's profile picture, Bob White, available, muted

5.2 Mobile Inspection Results

Our analysis of the use of the main functions in mobile versions of the three tools focused primarily on the effectiveness of the SR feedback, appropriate button positions and focus. We were unable to investigate specific gestures, as we would have liked to, since to the best of our knowledge none of the tools currently supports specific gestures for the main functions.

Analogously to Table 2 (desktop inspection), in Table 4, 'yes' is used to indicate (1) the focus is supported, (2) optimal positioning of the element (3) screen reader (SR) feedback is appropriate and a blank that the function is not available, inaccessible or not supported. The term "partial" is used to indicate a) the process to reach an object requires many steps or is complex b) performing a task or function is not intuitive or is too complex.

Table 4: Inspection evaluation of mobile Zoom

	Focus			Optimal Position			SR feedback		
	<i>Zoom</i>	<i>Meet</i>	<i>Teams</i>	<i>Zoom</i>	<i>Meet</i>	<i>Teams</i>	<i>Zoom</i>	<i>Meet</i>	<i>Teams</i>
F1. Joining a meeting							Yes	Yes	Yes
F2. Raising hand	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F3. Turning mic. On/off	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F4. Turning camera on/off	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F5. Screen sharing	Yes	Partial	Yes	Yes	No	No	Yes	Yes	Partial
F6. Using the chat	Partial	Partial	Partial	No	No	Partial	Partial	Yes	Partial
F7. Accessing shared content	No		No			Yes	No	No	No

F8. Checking mic&cam status	Partial	Yes	Partial	Yes	Yes	Partial	Yes	Partial	Yes
F9. Obtaining participant info	Yes	Partial	Partial	Partial	No	Partial	Partial	Yes	Partial

When carrying out different tasks via the SR VoiceOver on the smartphone, we observed the following interaction issues:

F1. Joining a meeting.

Joining a meeting is accessible via gestures for all the three tools. Joining Google Meet or Zoom by clicking on the link received by email or other messaging tools may raise some issues. The screen reader announces the names of people who join and leave the meeting. It can be very useful at the start of a meeting to know who is present. However, repeatedly hearing "Alex Smith has joined the meeting" or "Alex Smith has left the meeting", including when someone has briefly lost their internet connection, can quickly become irritating and distract attention and make it more difficult to participate in the meeting. Therefore, users should be able to customise these notifications through the application settings.

F2. Raising hand.

Raising a hand does not present any particular problems. The most significant issue is locating the button, which is not always easy to find by exploring the touchscreen.

F3 and F4. Turning microphone and camera. on/off.

The microphone and camera may need to be turned on and off several times in a meeting so being able to do this using a quick gesture would benefit users. The two buttons for turning the microphone and camera on/off are relatively easy to find for all three tools. This may be slightly more difficult when Zoom is used on a larger screen (e.g. iPad mini) as the buttons are located along the top edge and are sometimes confused with the speaker icons (announced by SR as avatars). In Google Meet, Teams and Zoom on smartphones the buttons are easier to find as they are placed along the bottom edge, an area where the user is used to looking for buttons. The screen reader clearly announces "mic. on" and "mic. off" when the button is clicked. This makes the SR feedback clear and easy to understand.

F5. Screen sharing.

This task presents no particular problems. However, there may be difficulties in exiting from this function in Google Meet or Teams. This could be due to the user not being aware that the sharing function called by the IOS system appears as a separate screen.

F6. Using the chat.

Chat messages are read automatically by the screen reader in Zoom but this could interfere with listening to the speaker or vice versa. Scrolling through chat messages to read them again requires the chat button to be located, which is not particularly easy in any of the three tools and therefore requires additional user interactions. When reading message in Google Meet, but not Teams and Zoom, the screen reader announces the writer and this could become irritating.

F7. Accessing shared content.

Content sharing is unfortunately inaccessible via screen reader in all three tools due to its graphical format.

F8. Checking microphone and camera status.

The microphone and camera on/off status can be checked in Zoom and Teams by reading the labels of the corresponding buttons and this requires the user to just place their finger over the button. The screen reader output is easy to understand. Having a "microphone on" or "camera off" sign in a screen corner or other position easy to locate with a finger could improve this function. This is the case for Google Meet, which makes this feature of the mobile version more usable than the desktop (web) application.

F9. Obtaining participant info.

Information about the status of other participants can help blind people decide whether to keep the camera on or off, including copying the behaviour of other participants. Information about microphone, but not camera, status is provided in all three tools. Camera status information is better supported in Zoom.

A brief overview of some of the features of gesture-based interaction with mobile devices will now be provided. Zoom has many good features and is, for instance, the only one of the three tools that can provide screen reader users with their camera status. The screen reader announcement of the speaker's name as soon as the microphone picks up the audio is very useful when a person is actually speaking. Unfortunately, this can be triggered by the slightest noise and the associated repeated announcement of names could quickly become irritating and distracting.

The mobile version of Google Meet is more usable than the (web) desktop version. The user interface is simple, though there is some difficulty in exiting from the sharing function. This is also the case in MS Teams. The chat is not automatically announced in Google Meet as is the case in Teams, but reading and editing messages is easy once the chat area has been opened. In Teams, when the chat is open new messages are read automatically.

5.3 Interacting via Screen Reader: two Examples

In this section we are presenting two examples to illustrate the keyboard steps that screen reader users have to take to obtain simple information. These cases relate to a conversation with another participant in the chat. The first case relates to reading this participant's messages and the second obtaining information about their status.

Case I: reading messages.

In this case a screen reader (SR) user is communicating in a chat window with another participant using the Teams tool. When the chat window is opened, the focus automatically moves to the edit field ready to write a message. To identify the last message from the other participant, the SR user has to press (1) Shift+Tab to move the focus to the message list; (2) up and down arrows to scroll through the messages and allow the screen reader to read them; (3) Tab to bring the focus back to the text box. This requires three steps. While this is not a very large number, it is multiplied each time the user has to do this and soon becomes large, so having to repeat these three steps for each incoming message is not very practical. The screen reader should (a) provide a command to read, for instance, the last (ALT+1), second (ALT+2) and third last messages (ALT+3) and (b) ensure that a message is read immediately after it arrives. This is not always possible for various reasons. For example, the SR user may be writing a message at the same time as receiving a message from the other participant. The SR cannot read out both the message the user is writing and the incoming message at the same time. Therefore the user will need to ask the screen reader to read the last message. This will happen automatically if the solution (a) is applied.

Case II: knowing another participant's status.

A SR user may want to know the status of another participant they are communicating with in the chat, for instance whether they are in the meeting, not in the meeting or giving a presentation. The video conference tool provides the user with this type of information in various ways, for instance by a colour code for their name or by showing textual information or icons next to their name. The screen reader user may need to use a number of different keyboard steps to bring the focus to the point where this information is reported. Further steps will then be required to return the focus back to the editing field.

We now consider how this works in Teams. We assume that the SR user has opened the chat in a window. The focus is then on the text box ready to write a message. Before sending a message, the user would like to know that the recipient is available and is not making a presentation. If, for instance, they are giving a presentation, there is a risk of the message

being seen by the audience when the screen is being shared. This type of information can be announced by the screen reader, for instance as 'available', 'communicating', 'presenting' (when screen sharing), 'absent' or 'offline'. Unfortunately, a large number of key presses is required to access the status which is shown next to the name in the list of chat names.

To obtain this type of status information, the screen reader user should (1) press the Tab key eight times or Shift+Tab nine times to bring the focus to the chat list of the names of the most recent chats. (2) press the up or down arrow keys or the 'read current line' command to read the name followed by their status (e.g., 'Alice, presenting', 'Bob, available', or 'Charlie, "away"'); (3) press the Shift+Tab eight times or Tab nine times to bring the focus back to the text box, or alternatively press the Enter key on the user's name as if activating it for the first time. It is not necessarily obvious to all users that this is what they are required to do.

While it is possible to carry out these keyboard actions, this requires a large number of steps. This is difficult for users and could lead to mistakes. In this case, the solution could be (a) a specific screen reader command that reads the status i.e. it carries out all the required steps without this being visible to the user or (b) adding the status immediately next to the participant's name e.g. 'Alice, "presenting" in the title of the chat window. In the latter case, the screen reader user only needs to use the key combination that allows them to read the title of the current window. This command is available in all screen readers.

6 END-USER SURVEY RESULTS

6.1 Participants

65 blind people participated in the desktop study and 94 in the mobile device one. Both studies were approximately male-female gender balanced with 31 females (47.7%) and 34 males (52.3%) in the desktop study and 45 females (47.9%) and 49 males (52.1%) in the mobile study. Six (6.4%) of the mobile participants and 10 (15.4%) of the desktop participants were partially sighted and the others were blind. All participants used screen readers and keyboards to access videoconferencing tools on both desktops and mobile devices. As figure 1 shows, both samples had an approximately bell-shaped age distribution with a longer tail to the left than the right and ages ranging from under 20 to 70+ years. However, the peak/mode was 40-49 years for the desktop group and 50-59 years for the mobile one. The relatively large number of participants over 70 for the mobile device, 12 (12.8%) is interesting, but further research would be required to determine whether older blind people prefer mobile devices to desktops.

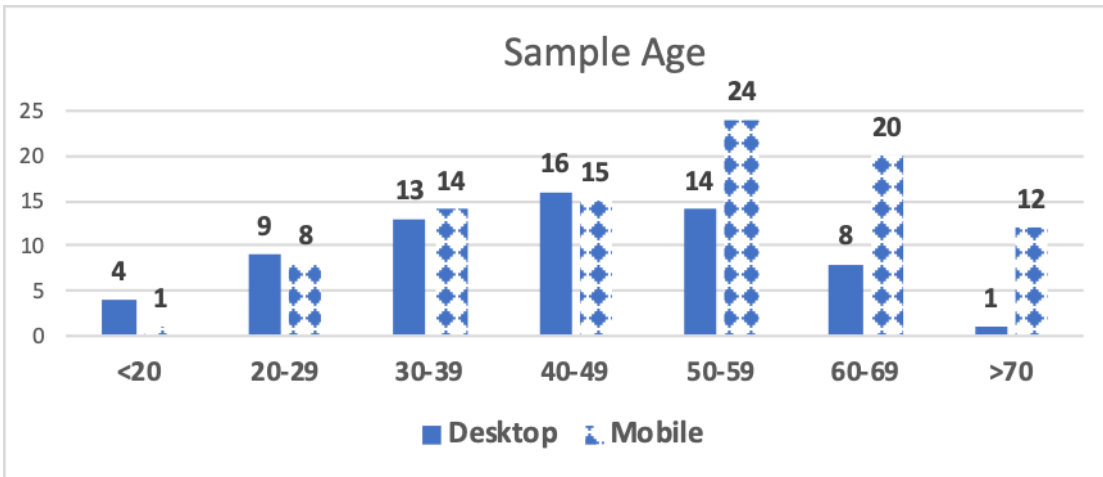


Figure 1: Age distribution

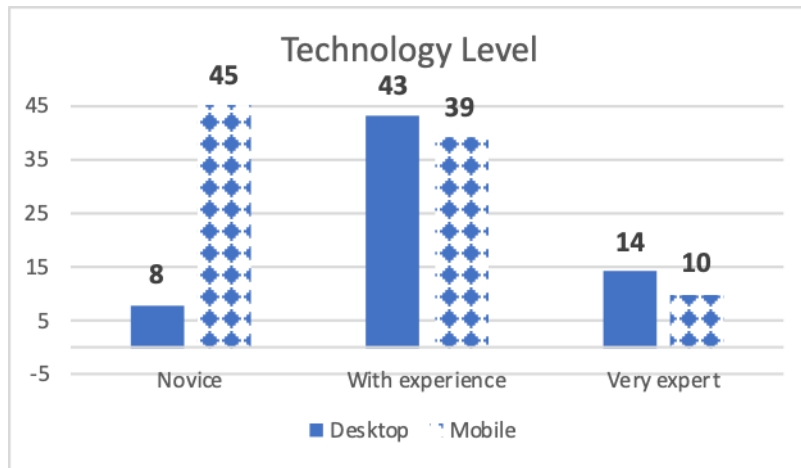


Figure 2: Technological skill

Based on their self-reports, as shown in figure 2, participants in both surveys had varied levels of technical experience and skills. Desktop participants were overall more experienced technology users than the mobile sample. They had nearly two thirds (66.2%) experienced users and just over a fifth (21.5%) very expert users with the remaining 12.3% novice users. Just under half (47.9%) of the mobile participants were novice users and just over two fifths (41.5%) experienced users with the remaining 10.6% expert or very expert users. The overwhelming majority of desktop participants (87.7%) used a PC with the Windows Operating System (OS) and the remaining 12.3% a Mac OS. The overwhelming majority of mobile participants (87.2%) used an iPhone. The remainder used an Android smartphone (4.3%) or an iPad (3.2%).

6.2 Accessing Video Conferencing Tools

Nearly all participants (96.8%) had used Zoom on mobile devices with close to 90% of them (87.3%) having used it on a desktop (87.7%). 80% had used Meet on a desktop and nearly as many, 78.7% on a mobile device. whereas only a

minority had experience of Teams, 15.4% on a desktop and 9.6% on a mobile device. Table 5 shows that for both desktop and mobile devices much higher percentages of participants had been able to access the basic functions of accessing a meeting and turning the microphone and camera on and off than were able to determine their own microphone or other participants' microphone and camera status or carry out less frequently used functions such as knowing who had raised their hand, screen sharing and file sharing. In general, a higher and sometimes much higher percentage of participants were able to access functions using Zoom than the other tools on both desktop and mobile devices, though this was not universally the case. A higher percentage of participants were able to access screen sharing and microphone status using Teams than Zoom on a desktop and camera and microphone status and the participant list using Teams on mobile devices.

The basic functions (i.e. turn on/off cam/mic, read/write chat and mic/cam status) were slightly more frequently used in Meet than in Teams, whereas more advanced functions such as Know participant mic/cam status, Write in chat and File sharing were more frequently used in Teams. Teams was mainly used by technologically skilled people i.e. 10 very expert, 20 experienced, and only one novice user in the desktop survey. This is in line with educational use in Italy where over the last two years many universities have adopted Teams video conferencing tools for lectures, meetings and exams and used its breakout features for laboratories and cooperative projects, whereas Google Meet was mainly used in primary and secondary education. This may have resulted in more experience with and greater ability to use the less frequently used functions or greater ability to overcome problems in using them. A lower percentage of participants was able to access most of the functions on each of the tools on mobile devices than desktop. However, the percentage of participants able to access file sharing on Zoom on mobile devices was twice that on desktops and slightly greater on Teams and the percentages for raising and lowering a hand were all greater on mobile than desktop.

Table 5: Percentages of participants able to access the different functions

	Desktop			Mobile Device		
	Zoom (n=91)	Meet (n=75)	Teams (n=9)	Zoom (n=57)	Meet (n=52)	Teams (n=31)
Accessing a meeting	100	100	97	100	99	100
Turning mic on/off	100	100	100	100	100	100
Turning camera on/off	95	98	97	93	81	56
Other participants' mic status	37	12	58	8	4	44
Other participants' cam status	35	12	26	4	3	22
Knowing your mic status	88	83	81	27	21	56
Knowing your cam status	86	83	84	21	15	56
Raising/lowering hand	88	38	68	90	71	78
Knowing who has raised a hand	30	10	29	20	11	33
Writing in chat	79	58	61	90	88	100
Reading chat	81	50	58	58	39	67
Accessing participant list	86	48	68	52	31	67
Screen sharing	54	48	61	31	8	11
File sharing	19	12	10	44	19	11

Table 6 indicates which differences in the percentages of participants able to access particular functions on desktop and mobile platforms in table 5 are statistically significant for each of the three tools, Zoom, Teams and Meet by giving the χ^2 and p values. For example, the greater percentage of Desktop Zoom users (81%) who were able to read the chat compared to Mobile Zoom users (58%), was found to be statistically significant ($p = 0.004722$), whereas the greater percentage for Desktop Meet users (50%) compared to that for Mobile Meet users (39%) was not statistically significant

(0.20498). From table 6 over two fifths (43%) of the differences between the percentages of participants able to access particular functions on desktop and mobile devices are statistically significant at the 0.05 level. The table does not include the chi squared and p values for accessing a meeting and turning the microphone on and off, as these percentages are very close to each other and to 100%. The values for Teams for writing in the chat have been modified slightly in the calculation to remove a non-zero value. The differences are statistically significant for all three tools for only one function, screen sharing. Zoom and Meet both have eight statistically significant differences, with only half of them for the same functions, and Teams only two.

Table 6: Statistical significance of differences in percentages of participants able to access functions on desktop and mobile devices

	Zoom (χ^2 , p)	Meet (χ^2 , p)	Teams (χ^2 , p)
Turn cam on/off	0.109, 0.741756	8.27, 0.004041	10.83, 0.000996
Raise/lower hand	0.208, 0.648554	13.03, 0.000306	0.33, 0.563005
Know who raised hand	1.96, 0.161698	0.04, 0.847594	0.06, 0.804228
Access to participant list	18.11, 0.000021	3.96, 0.046615	0.004, 0.951653
Screen sharing	8.15, 0.004296	26.73, < 0.00001	7.03, 0.008038
Read the chat	7.98, 0.004722	1.61, 0.20498	0.22, 0.642835
Write in chat	3.59, 0.058248	15.29, 0.000092	2.88, 0.089794*
File sharing	9.44, 0.002129	1.18, 0.278168	0.016, 0.899563
Know my mic status	50.89, < 0.00001	38.98, < 0.00001	2.34, 0.125952
Know my cam status	59.78, < 0.00001	58.14, < 0.00001	3.21, .073322
Know participant mic status	19.41, .000011	2.65, 0.103525	0.52, 0.469654
Know participant cam status	24.30, < 0.00001	4.10, 0.04301	0.05, 0.826955

6.2 Exploration Strategies

As shown in figure 3, shortcut keys were found to be useful, but not used universally. The majority of participants used shortcuts either sometimes (63.6%) or always (32.7%) on Zoom, and sometimes (73.3%) or always (20.0%) on Teams with lower percentages on Meet: 56.5% sometimes and 13.0% always. Over a fifth (30.4%) never used keyboard shortcuts on Meet. This raises the issue of why keyboard shortcut keys were so relatively popular on Zoom and so unpopular on Meet. Determining the reasons for this will require further research, but they could include Zoom having shortcut keys for more functions, greater familiarity with Zoom and consequently greater awareness of the shortcut keys. The results on ease of navigation without shortcut keys in the next section show that the explanation is not greater difficulties in using Zoom than Meet or Teams without shortcut keys.

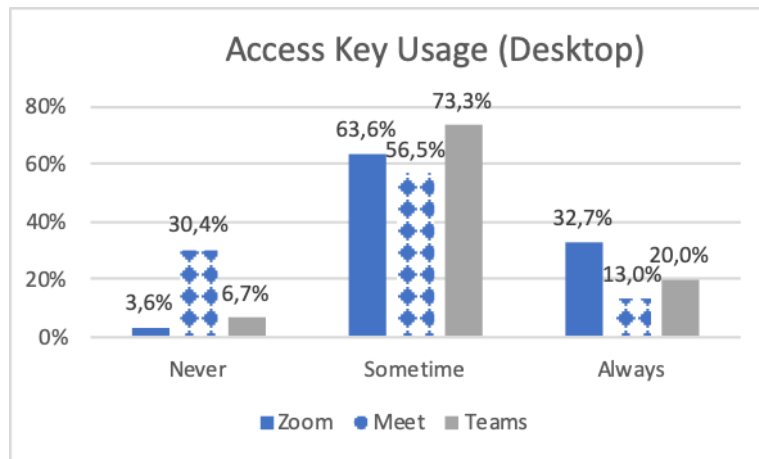


Figure 3: Use of Shortcuts with Zoom (left bar) N=55, Meet (middle patterned bar) N=46 and Teams (right bar) N=30

As shown in figure 4 left and right flick gestures were the main strategy used for exploring the smartphone screen (user interface) on all three video conferencing tools. 90% of participants used gestures on Meet and 86% on Zoom, but only two thirds on Teams. The strategy of putting a finger in a precise screen position was a very distant second used by 9% of Zoom and Meet participants, but 22% of Teams ones. Only a few participants explored the whole screen for all three tools.

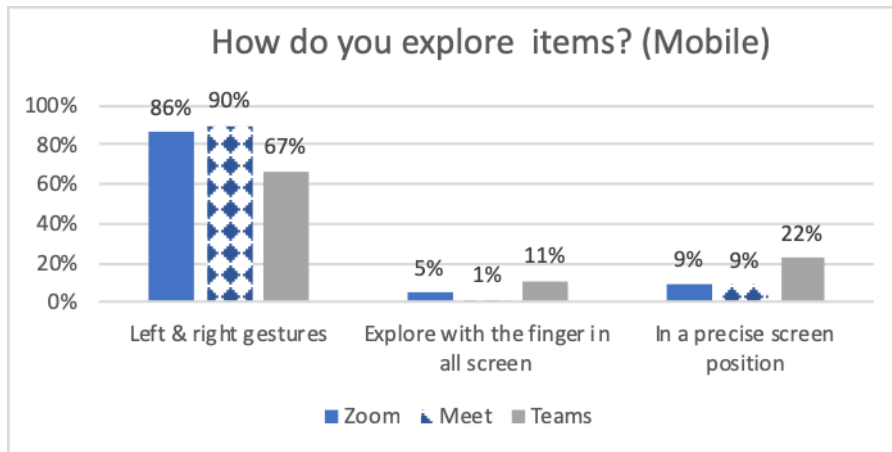


Figure 4: UI exploration strategies for Zoom (N=86), Meet (N=69) and Teams (N=9). Left&right gesture (left line-patterned bar), finger exploration of the whole screen (middle bar chart) and a precise screen position (right point-patterned bar)

6.3 Ease of Use and Satisfaction

Figure 5 shows that users did not find it particularly easy to navigate without shortcuts and that a relatively high percentage, 50.9% on Zoom and 65.2% on Meet, found this neither easy nor difficult. A higher percentage of participants found Zoom easy or very easy to navigate than Meet or Teams, though the value was still low at only 30.9%. Teams was

found to be the most difficult to navigate with 70.0% finding it difficult or very difficult to navigate. This could be due to less familiarity with Teams than the other two tools, but there may also be aspects of Zoom’s structure which make it easier to navigate. However, a stronger conclusion is the importance for keyboard shortcuts for easy navigation.

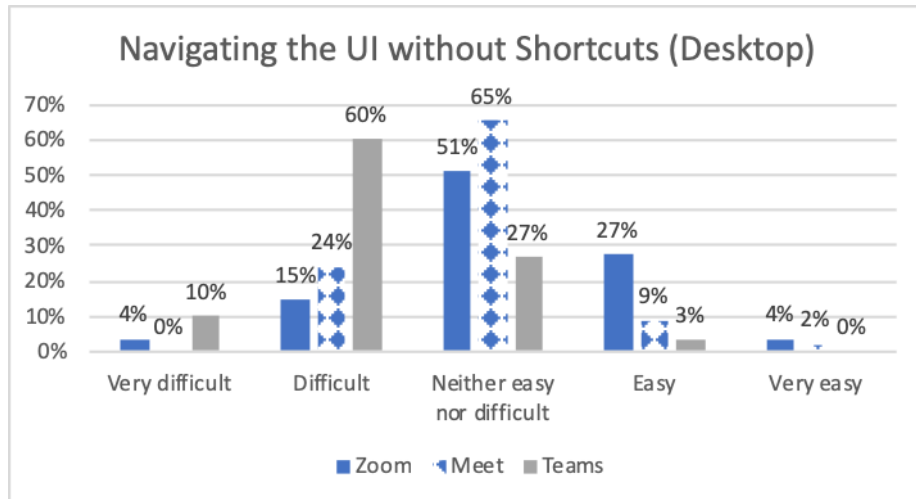


Figure 5: Ease of navigation without keyboard shortcuts. Zoom (left bar) N=55, Meet (middle patterned bar) N=46 and Teams (right bar) N=30

On mobile devices, in contrast to the desktop results, Teams was evaluated as easy to use (easy and very easy) by 57% of participants and Zoom and Meet as easy or very easy to use by only 25%, whereas about two thirds answered “I do not know”. However few of the participants were advanced users of Teams on mobile devices, so the comparison may have included both basic and advanced functions on Zoom and Meet and only basic functions on Teams. In addition, the small number of users of Teams also complicates comparisons.

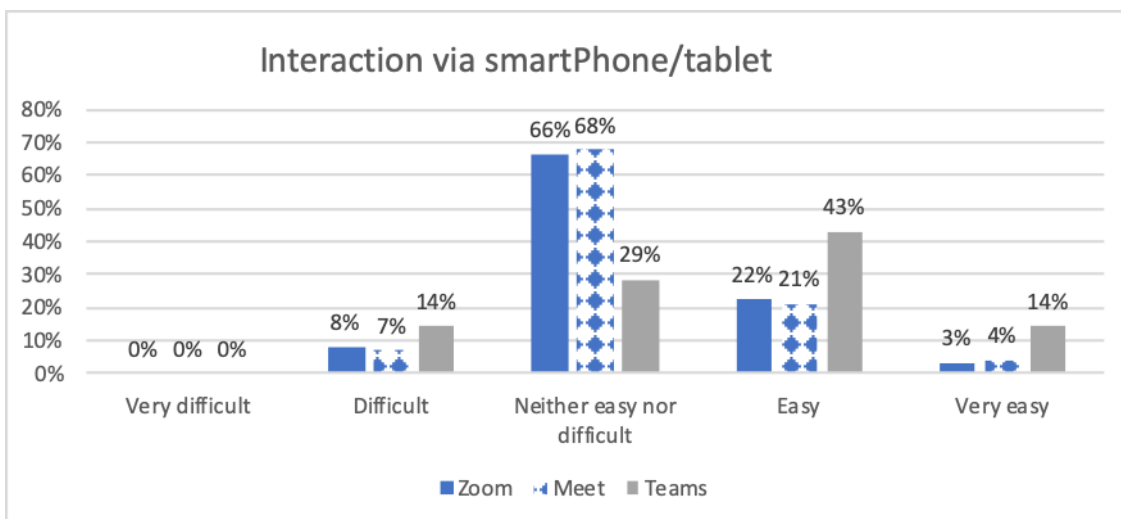


Figure 6: Ease of use of Zoom (left bar) N=87, Meet (middle patterned bar) N=71 and Teams (right bar) N=7 on mobile devices

Concerning user satisfaction, Figures 7 and 8 show that participants had the greatest satisfaction with Zoom on both desktop and mobile devices with two thirds (81.5%) satisfied or very satisfied with it on desktops and 73.8% on mobile devices. This was followed by Teams on desktops with just over a half (56.7%) satisfied or very satisfied, and Meet on mobile devices with just 37.5% satisfied or very satisfied. Few participants were (very) dissatisfied with any of the tools on mobile devices, indicating that many of them did not answer the question. On desktops more than a quarter, (27.6%) were dissatisfied with Meet and 16,7 % with Teams vs Zoom only 1,9%.

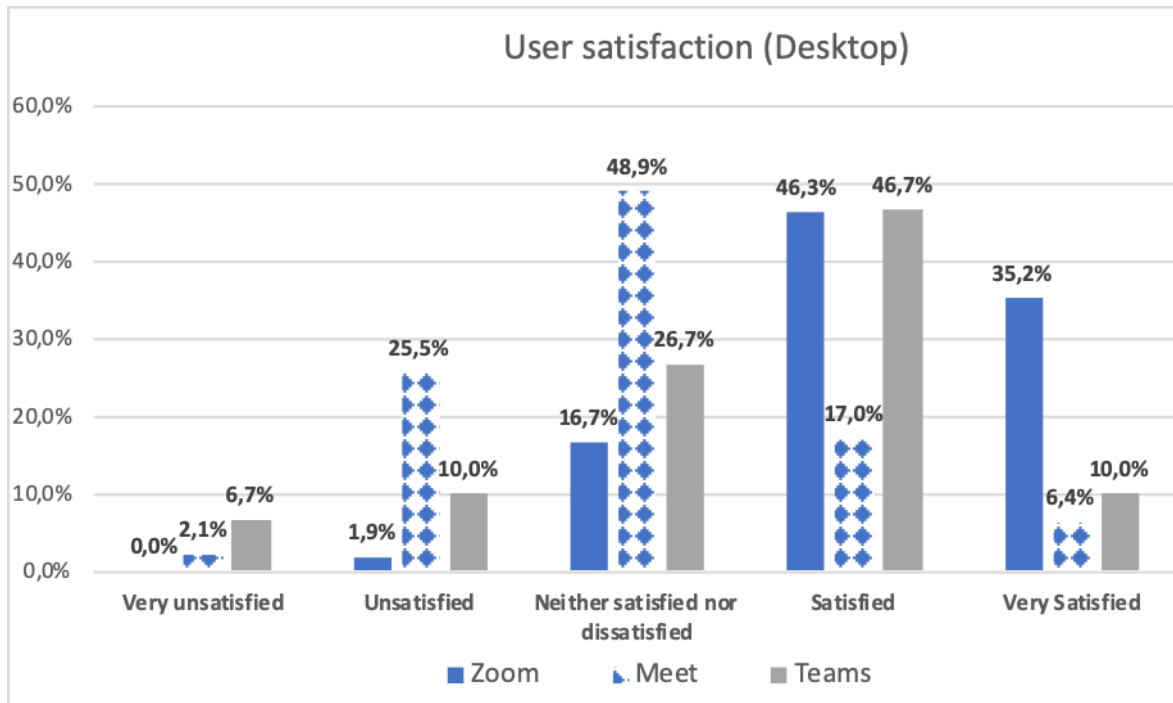


Figure 7: User satisfaction with Zoom (left bar) N=54, Meet (middle patterned bar) N=47 and Teams (right bar) N=30

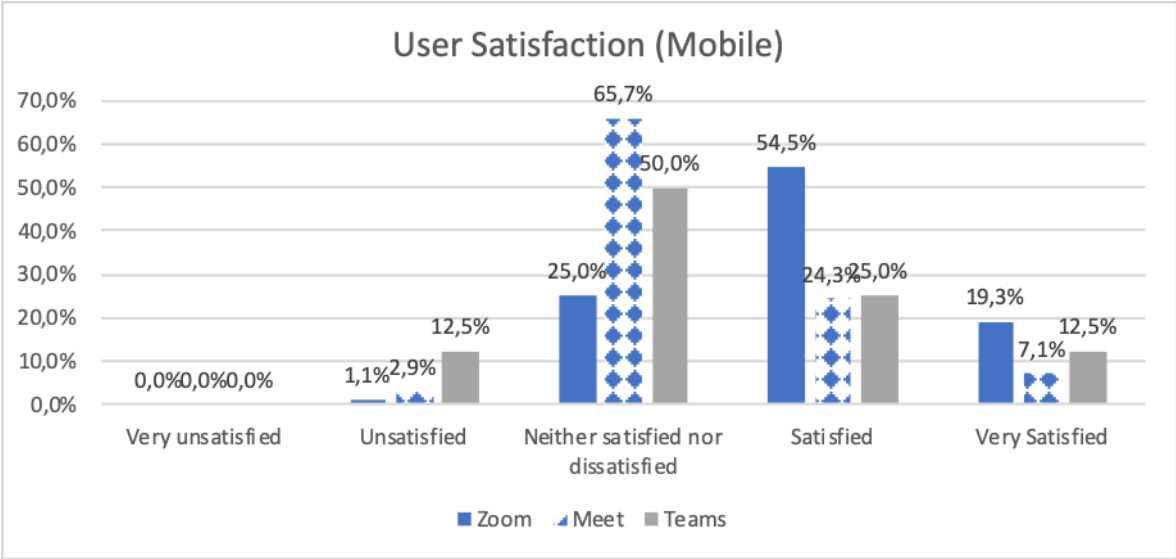


Figure 8: User satisfaction with Zoom (left bar) N=88, Meet (middle patterned bar) N=70 and Teams (right bar) N=8 on mobile devices

Figure 9 shows that participants did not find automatically read content particularly useful. Few participants considered this (very) useful, only 12% on Zoom, 17.2% on Meet and 25% on Teams. Determining why this was found to be more useful on Teams than the other tools would require further investigation. However, we have to consider that in our sample only few participants used Teams compared to Zoom and Teams. Participant comments indicated that, for instance, ‘It is irritating that the screen readers read out everyone who enters and leaves’. This participant also noted ‘you can configure it in Zoom with a script but not in Meet’. However, this could be difficult for users who were not particularly experts.

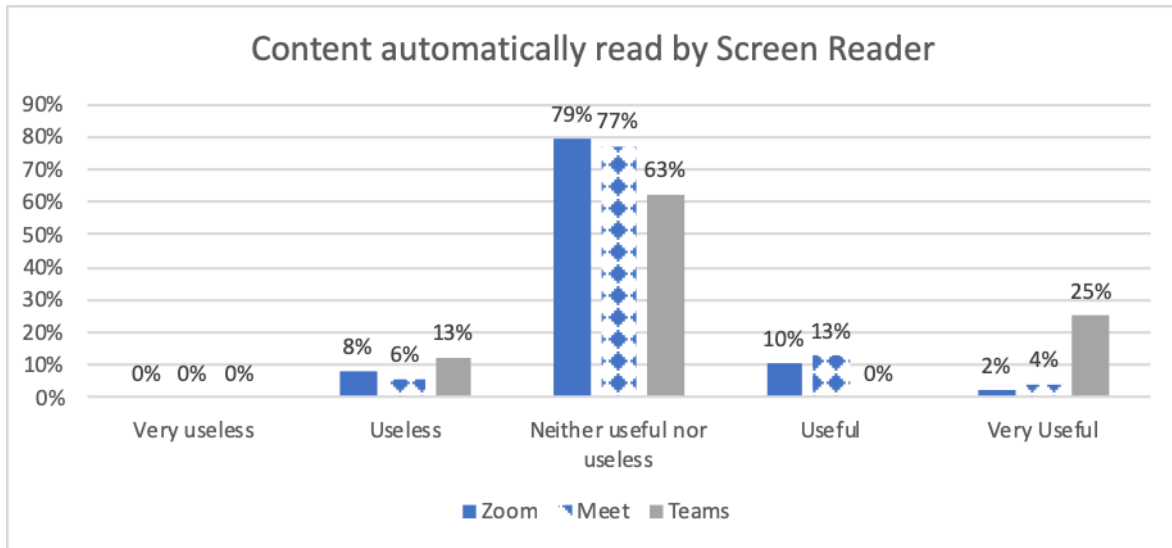


Figure 9: Usefulness of content read automatically by SR: Zoom (left bar) N=87, Meet (middle patterned bar) N=70 and Teams (right bar) N=8

Concerning the preferred access devices, participants had a strong preference for using Zoom (86.2%) and to a lesser extent Meet (75.3) with an app on a computer (see figure 10). Phone/tablet was a distant second, preferred by only 12.6% on Zoom and 24.7% on Meet. Only small numbers answered this question for Teams, with twice as many preferring an app on a computer to either phone/tablet or web. The difference in preferences between the two platforms was lowest for Meet which offers an “accessible” web version for both desktop and mobile devices. However, participants were probably not aware that the computer version of Meet is actually a Web application.

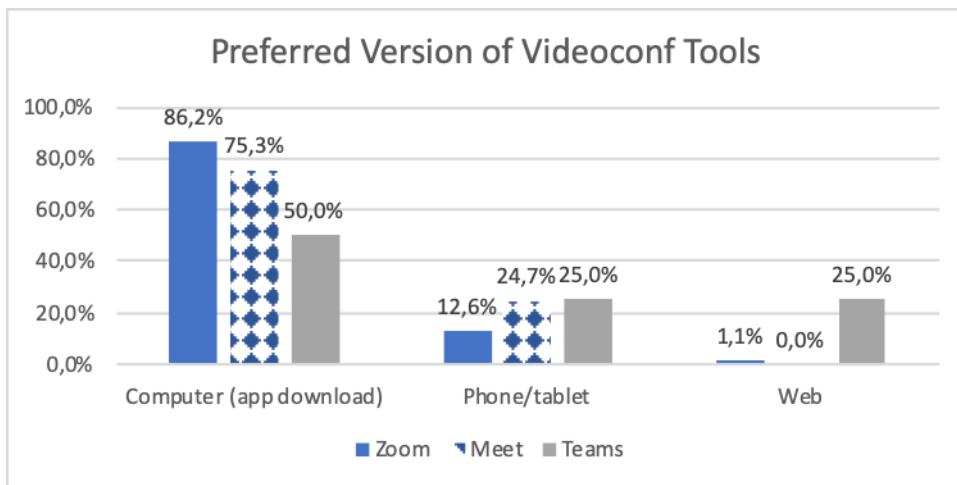


Figure 10: Preference for using tools on computer, phone/tablet and web: Zoom (left bar) N=87, Meet (middle patterned bar) N=73 and Teams (right bar) N=8

6.4 Participant comments

Despite the much greater number of responses to the mobile devices than desktop survey, over four times as many participants commented on the tools on desktops (44 or 68%) than mobile devices (8 or 8.5%). Probably unsurprisingly, they generally commented on problems rather than features they were satisfied with. They also suggested additional functions they would find useful and commented on which tools they used, sometimes with explanations or indications of how they used them. For instance, comments from desktop participants included, 'I only use Zoom ... I am not familiar with the others' and 'I use Zoom in the association and Teams and Meet at university' (mobile participants) and 'I use Google at work particularly when I work from home. I've used Zoom with JAWS for some online courses' and 'I use Zoom. I have tried Google only a few times and not tried the Microsoft tools at all.'

One of the most frequently raised problems with tools on desktop devices related to the complexity of the user interface (20 participants). They considered it to have too many elements, to be difficult to navigate and time consuming. For instance, 'these tools are hellish. I never understand anything, where I am and what I need to do ... Fortunately there are shortcut keys.' Another significant problem was the lack of shortcuts (24 participants) so that functions required a lot of key presses to carry out. For instance, 'It is not easy to go from one area to another in the Teams interface in a meeting ... unless you use shortcuts'; "improve the interaction with shortcut keys to know the numbers of raised hands in Zoom" and 'I would like to be able to read shared content when it is presented, move quickly between the chat and the participant list and see the state of the microphone and video camera without needing to turn them on and off and use a lot of buttons to see whether they are on or off'. There was also a suggestion to standardise the shortcut keys to avoid 'confusion between platforms'.

Six participants considered the interface too difficult to learn and, sometimes, to require assistance from a sighted person. For instance, 'I need to use the tools for a bit at the start, as it's difficult to understand them initially, particularly Teams. I had to ask someone to explain how the interface is organised.' Examples of the resulting difficulties included "knowing who is speaking is a great problem" and "rereading the last message in the chat needs too many Tab and Arrow keys". These issues are clearly related, as additional shortcuts would reduce interface complexity to some extent and navigation difficulties, lack of shortcuts and interface complexity make the tools more difficult to learn. These comments are in line with the quantitative results which indicated relatively frequent use of shortcuts, particularly on Zoom, and some difficulties in navigating without them. Copying links or content from the chat also caused difficulties, for instance, 'I still find it difficult to copy a link or a written text (e.g. if a phone number is written down and I want to copy it)'. There was a suggestion of 'a function which quickly visualises links and documents and distinguished them from other chat messages'. Three participants commented on the need for flexibility "to personalise and simplify the interface", for instance, "to visualise the things you need".

Participants experienced difficulties with the focus, losing focus and orientation in the user interface on desktops and having to bring the focus back to the top of the page after the first screen, as 'it does not position itself automatically' on mobile devices. "I often lose messages in Google Meet as the screen reader loses the focus or cuts off messages." Interference between the meeting speakers and screen reader information could also cause problems. Desktop participants found the announcements of the names of participants 'joining and leaving Zoom irritating' and that the repeated announcements of participants as speakers when the microphone detected background noise could lead to loss of concentration. However, mobile device participants preferred the 'clear' announcements of 'who was speaking' on Teams to the need to 'deduce' this on Meet and Zoom. Comments on tool use on mobile devices also indicated they preferred the computer versions due to 'better control of the camera' and the ability to 'take notes ... while listening', whereas it was 'awkward' to listen to a lecture on one device and take notes on another. They preferred 'to do everything

on the computer’. Mobile device users also complained about having to type the link into an app when joining and one participant was denied access to Teams.

Desktop participants were particularly interested in additional shortcuts to simplify and speed up their interactions (24 participants) e.g. ‘I am suggesting more keyboard shortcuts .. everything should be faster and immediate’ They considered shortcuts could provide fast access to grouped commands to simplify the user interface (20 participants). .In particular, they wanted shortcuts for (1) the list and number of participants, (2) list of chat messages, (3) files available to participants, (4) their microphone and camera status and the status of the virtual background, and (5) the list of participants with cameras on. Specific comments about the need for additional shortcuts included ‘A command to know whether the video, microphone and background are on or off would be useful. I would also like a command to know who is speaking’ and ‘It would be useful to have some more keys, for instance to quickly have the list of participants, know whether the mic/video are active, raise a hand’.

There was some overlap between desktop and mobile device users in their desire for shortcuts, as the latter were also interested in functions for rapidly knowing the speaker in Meet and Zoom, receiving more feedback and improving focus handling. Both groups of users wanted the ability to control or silence the screen reader for some functions or announcements with mobile device users wanting the option to deactivate the announcement of names and people arriving and leaving. However, two desktop users wanted to be aware of the context and ‘the screen reader to read more things to enable the user to understand where they are, otherwise it is too difficult to listen to a meeting and understand what is on the screen at the same time’. Finally, six desktop users wanted tutorials, user interface commands and a description of the interface to facilitate and speed up learning it e.g. ‘It would be useful to have tutorials which explain how to use video conferencing including with a screen reader’.

7 DISCUSSION

65 blind people responded to the survey on videoconferencing tools on desktops and 94 to the version for mobile devices. Comparison of the survey results and those of the inspection evaluation for desktop and mobile devices showed full agreement on the main issues.

Analysis of the results showed that all three tools were generally able to provide means of accessing their main functions via keyboard or gestures with screen reader feedback. However, access was frequently not easy due to the lack of shortcuts (desktop) and the lack of specific gestures (mobile) for the main functions and other problems resulting from poor design for screen reader feedback. Clear and unambiguous feedback enables users to interact with the system and monitor the status of other users in the collaborative online environment [14]. In the desktop version, Zoom was found to be the most accessible of the three tools, probably due to the numerous shortcuts available for the main functions, whereas Google Meet was preferred in the mobile version, possibly due to having a simpler interaction with the user interface than Zoom. However, all tool preferences on mobile devices were low. Table 7 shows a summary of the accessibility features supported by desktop and mobile video conferencing tools, as determined by the inspection evaluation of the user interface.

Table 7: Summary of accessibility features supported by the desktop and mobile video conferencing tools

<i>Function</i>	Desktop			Mobile		
	<i>Zoom</i>	<i>Meet</i>	<i>Teams</i>	<i>Zoom</i>	<i>Meet</i>	<i>Teams</i>

F1. Joining a meeting	Yes	Yes	Partial	Yes	Yes	Yes
F2. Raising hand	Yes	Partial	Partial	Yes	Yes	Yes
F3. Turning mic. on/off	Yes	Yes	Yes	Yes	Yes	Yes
F4. Turning camera on/off	Yes	Yes	Yes	Yes	Yes	Yes
F5. Screen sharing	Yes	Partial	Yes	Yes	Partial	Partial
F6. Using the chat	Yes	Partial	Yes	Partial	Partial	Partial
F7. Accessing shared content			Partial	No	No	No
F8. Checking mic. and camera status	Yes	Yes	Partial	Partial	Yes	Partial
F9. Obtaining participant info	Yes	Partial	Partial	Partial	Partial	Partial

The study highlighted that none of the three tools was fully accessible via screen reader and keyboard in the desktop environment and gestures in the mobile environment. The overall preference was for Zoom on desktop. Both the questionnaire and inspection evaluation found that it was able to support the greatest number of basic functions required for participation in an online meeting. However, users reported that they were unable to use several of the functions that the inspection evaluation determined could be accessed by keyboard and screen reader, but which required a non-linear or lengthy procedure. Examples include checking other participants' status (e.g. microphone, video camera and raised hands) and accessing and using the chat.

This shows the importance of end-user testing and the need for both accessibility and usability so that functions are not just theoretically accessible, but easy and intuitive to use in practice, including by non-expert users. Long and complex procedures are likely to act as a barrier, particularly for non-expert users. User feedback was important in identifying usability issues and the inspection evaluation in explaining them. Making all functions easy and intuitive for all users makes it easier for them to focus on the meeting, what is being said and any contributions they may want to make rather than their attention being distracted and energy dissipated by difficulties in using the tools. This is particularly important for screen reader and keyboard users with minimal experience with the tools.

Comments from several users indicated they were unable to use some of the functions. This may be due to a lack of easily available information on how to use them or usability issues. Inspection evaluation showed that some information, for instance on video or microphone status, could be read next to the participant's name in the participant list. However, when there was a lot of information and each line was lengthy, participants could experience difficulties in accessing and understanding the displayed information, making it not very useful in practice. In addition, as indicated in the results section, it was not always easy to access the participant list.

Searching for specific information, particularly during a meeting or trying to access the list of chat messages could require considerable cognitive effort, as the user tried to understand both the speaker(s) and the information being read by the screen reader. Non-disabled users avoid these problems by using different sensory modalities to search for information and listen to a meeting, whereas screen reader users use hearing for both. This risks cognitive overload and interference and the screen reader information making it difficult to understand the speakers or vice versa. This makes it essential that all information is easily available through a minimal number of simple steps, preferably just one, and complex mechanisms involving multiple steps are not required.

Some participants found it very difficult to join Teams meetings on a desktop, as the "participation" button is not clearly visible and several steps were required to get to it. Some participants experienced difficulties in clicking on their

invitation link in Zoom and Meet. In Teams shared content was partially accessible in the desktop version as a ppt file, whereas it was only possible to read the number of slides but not access the content in the mobile version.

The results of the desktop survey largely confirm those of the earlier study by Leporini et al. [39] of which they form an extension. There are small differences, due to the larger sample (65 rather than 29). For reader convenience, the results of the initial study are reported in Table 8 while table 9 shows user preferences for the desktop video conferencing tools in this extended version. Two five-point Likert scales were used for measuring values of ease of use (on the left) and user satisfaction (on the right).

Table 8: User preferences for the initial study of desktop video conference tools (29 participants) [39]

DESKTOP (29 users)	Ease of Use			User Satisfaction		
	Zoom	Meet	Teams	Zoom	Meet	Teams
M	3,34	2,71	2,27	4,14	3,07	3,14
SD	0,72	0,60	0,70	0,69	0,65	1,10

Table 9: User preferences for the desktop video conferencing tools (65 participants)

DESKTOP (65 users)	Ease of Use			User Satisfaction		
	Zoom	Meet	Teams	Zoom	Meet	Teams
M	3,13	2,89	2,23	4,15	3,07	3,43
SD	0,84	0,64	0,68	0,69	0,65	1,10

In the desktop environment, participants considered interaction and navigation with Meet and Teams not to be easy and with Zoom to be moderately easy/difficult. On a five-point Likert scale with 5 indicating greatest ease of use, mean values (M) and standard deviations (SD) were Zoom (M=3.13, SD=0.84), Google Meet (M=2.89, SD=0.64), and Teams (M=2.23; SD=0.68). A 5-point Likert scale evaluation with 5 indicating greatest satisfaction, showed that overall participants preferred Zoom (M=4.15, SD=0.69), followed by Teams (M=3.43; SD=1.10) and Meet (M=3.07, SD=0.65). Participants had a much greater range of opinions about Teams than Zoom and Meet, giving a much higher standard deviation (1.10 compared to 0.69 and 0.70 respectively).

Analogously Table 10 shows user preferences for the mobile video conferencing tools.

Table 10: User preferences for the mobile video conferencing tools with 94 participants

MOBILE	Ease of Use			User Satisfaction		
	Zoom	Meet	Teams	Zoom	Meet	Teams
M	3,21	3,23	3,57	3,92	3,36	3,38
SD	0,63	0,64	0,98	0,70	0,66	0,92

A five-point Likert scale of tool use on mobile devices with 5 indicating greatest ease of use, mean values (M) and standard deviations (SD) found that Teams was the easiest to use (M=3.57; SD=0.98) followed by Meet (M=3.23, SD=0.64), and Zoom (M=3.21, SD=0.63). However, the greater ease of use of Teams may have been due to responses from only a small number of mainly expert or very expert users.

A 5-point Likert scale evaluation with 5 indicating greatest satisfaction, showed that overall participants preferred Zoom (M=3.92, SD=0.70), followed by Teams (M=3.38; SD=0.92) and Meet (M=3.36, SD=0.70). Participants seem to have found the interaction via gestures on mobile devices to be simpler than keyboard interaction on the desktop version.

However, overall they preferred the desktop versions, probably due to the flexibility offered by the desktop environment. Participants used these tools at university and at work. They needed to take notes, share content, and move between applications such as email and this can be done more rapidly and easily using a keyboard in a well known environment.

Table 11 shows that the differences in the values for Zoom, Meet and Teams are statistically significant at the 0.05 (and higher) levels for ease of use for both desktop devices for both samples, but not mobile devices. These differences are significant at the 0.05 (and higher) levels for satisfaction for both desktop and mobile devices.

Table 11: Table (χ^2 , p) for differences between values for Zoom, Meet and Teams

	Ease of use	Satisfaction
Desktop, n=29	(23.045, < 0.00001)	(26.49, 0.000025)
Desktop, n=65	(25.33, 0.000043)	(32.45, 0.000013)
Mobile, n=94	(4.84, 0.564249)	(23.53, 0.000013)

As shown in table 12, the differences in the values for desktop and mobile devices are only significant for both ease of use and satisfaction for Meet. The lack of significance for ease of use for Zoom and satisfaction for Teams is unsurprising due to the closeness of the means. The small number of respondents for mobile devices for ease of use for Teams may have affected the result.

Table 12: Table (χ^2 , p) for differences between values for desktop (n=65) and mobile (n=94)

	Ease of use	Satisfaction
Zoom	(3.41, 0.33)	(2.89, 0.41)
Meet	(14.08, 0.0028)	(11.11, 0.011157)
Teams	(2.20, 0.33)	(1.58, 0.66)

Tables from 13 to 18 show ease of use and satisfaction for Zoom, Meet, Teams, as expressed by users.

Table 13: Zoom: Ease of use. DT = desktop, Mob = mobile device

Zoom	Ease of Use					Mean
	1	2	3	4	5	
DT	2	10	45	23	4	3.19
Mob	0	7	59	20	3	3.21

Table 14: Zoom: Satisfaction

Zoom	Satisfaction					Mean
	1	2	3	4	5	
DT	0	1	14	39	23	4.21
Mob	0	1	22	48	17	3.92

Table 15: Meet: Ease of use

Meet	Ease of Use				
------	-------------	--	--	--	--

	1	2	3	4	5	Mean
DT	2	15	49	4	1	2.83
Mob	0	5	44	15	3	3.23

Table 16 : Meet: Satisfaction

Meet	Satisfaction					Mean
	1	2	3	4	5	
DT	1	13	42	10	4	3.02
Mob	0	2	46	17	5	3.36

Table 17: Teams: Ease of use

Teams	Ease of Use					Mean
	1	2	3	4	5	
DT	4	24	11	2	0	2.24
Mob	0	1	2	3	1	3.57

Table 18: Teams: Satisfaction

Teams	Satisfaction					Mean
	1	2	3	4	5	
DT	3	6	12	15	3	3.34
Mob	0	1	4	2	1	3.38

Overall the study indicated the importance of usability as well as accessibility. While many of the available functions could be used, particularly on Zoom, user experiences could be considerably improved and many functions made much easier to use.

Microphone and camera handling are well supported by the three tools on both platforms. However, camera control could be improved by providing information about background images and blur and helping blind participants to correctly focus their cameras on their faces. Support for camera framing would be useful and could be provided by artificial intelligence algorithms, as is already offered for instance by the camera in the Apple IOS for face centering. Access to the chat could be improved by making messages easier to read and not reading out additional content such as "Bob says..." or "Alice answered in the conversation...", as currently occurs in some Tools such as Google Meet and Teams. Messages should include only the essential content. The user interface should be organized to minimise the number of actions required to operate a function and explore the resulting content. For example, it should be possible to scroll through participants' names in one or at most two steps.

One of the difficulties in using video conferencing tools via screen reader results from the fact that all information is available through the audio channel. Consequently, users frequently need to listen to the screen reader for tool information and the speaker at the same time. This requires a lot of cognitive effort from them, again stressing the importance of minimising the number of steps required to access each function.

Many other such suggestions could be provided. It should also be noted that the problems identified are due to the tools not being designed to take account of screen reader functionality and limitations.

8 GUIDELINES FOR DEVELOPERS

The guidelines in this section draw on the results of the study with the implementation details developed by the researchers to most effectively implement them. They are intended to support developers of video conferencing software and assistive technologies in improving accessibility and usability for screen reader users. They go beyond the W3C-WAI draft guidance for software for remote meetings [63] in providing detailed guidelines related to the user interface for developers of video conferencing technologies for screen reader users.

8.1 Principles

The guidelines are based on the following principles which have emerged from the research:

1. Providing easy access to the same information as non-disabled users while minimising cognitive load.
2. Providing easy to implement customisation options which do not overcomplicate the system.
3. Easy access to functions and information through shortcuts (desktop), gestures (mobile devices) or the Tab key.
4. Clear audio of both speakers, and messages and feedback from the screen reader, with the means to minimise interference between them.

8.2 Guidelines

Guidelines G1 and G2 are part of a three step approach to making commands and functions easily accessible via shortcuts or gestures. This involves (a) organizing the interface into functional areas or panels; (b) making each area (or panel) accessible via keyboard or gesture; and (c) moving the system focus to the area. Step (a) is implemented through G1 and (b) and (c) through G2.

G1. Partitioning the user interface to arrange and group functions and commands logically related.

Developing areas and panels (tabs) that group functions logically simplifies the interaction and the search for functions and information. Panels in the user interface can be used to group video calls, participants and chat functions. Functions could be further grouped by type on toolbars, menus or frames in each area or panel. This logical structure will make it easier for users to remember where functions are and easily find them with a small number of key presses (G2). This will reduce cognitive load and enable users to focus on the meeting rather than looking for functions.

1. Developing common blocks e.g. areas, panels, tabs or separate windows to group functions and information by type e.g. buttons to switch microphone and camera on and off, share the screen and access chat messages.
2. Placing the most commonly used function and command buttons in easy-to-find locations, e.g. the microphone and camera on/off buttons near the bottom (or top) edge, makes it easier and quicker for users to find them, particularly on a touchscreen.

G2. Making it easy to locate the focus on all UI blocks, functions and commands and operate them via the keyboard or gestures in an effective way

The aim of this guideline is to ensure that the focus can be moved from one block to another using keyboard shortcuts or the tab key (desktop) and gestures (mobile devices) and the keyboard or gestures can be used to interact with elements of the block the focus is on. Specific design recommendations include:

1. Basic operability via keyboard e.g. with focus handling via the Tab, Ctrl+Tab and Arrows keys.
2. Basic operability via gestures, e.g. Three-finger swipe right and left to skip from a panel to the next/previous one.

3. Assigning specific shortcuts to each (frequently used) block, panel or toolbar to make it easier and quicker to access them and reduce the number of key presses required.
4. Assigning specific shortcuts or gestures to frequently used functions and commands to make it easier and quicker to access them and reduce the number of key presses required.
5. Automatically moving the focus onto a block or panel when the shortcut or gesture for this block or panel is pressed.

G3. Providing appropriate screen reader feedback on the status of input and output devices, participant information and chat content.

Since screen reader users do not have direct access to what is on the screen they require SR feedback to inform them when a function is triggered or the screen changes. Specific design recommendations include:

1. Feedback should be short and relevant e.g. short sound or very brief message to indicate an action has taken place, and avoid redundant and superfluous comments and content.
2. SR status information about input devices or other events should be easily available through shortcuts or gestures or automatically when a device is triggered. It should be very clear and not require interpretation e.g. “mic on” or “microphone is on” not “switch the microphone off”.

G4. Providing SR information about the speaker and the content presented.

SR users do not automatically have access to information about the speaker or presenter and the title of any presentations, whereas this information is generally available to non-disabled people from the screen.

1. The SR should automatically or on demand provide the name of the person currently speaking or presenting (e.g. slides or videos) in a format chosen by the user.
2. To avoid repeated announcements of the speakers’ names being triggered by background sounds or muttering, only sounds recognised as speech should be included (requiring voice recognition software) and a threshold volume used.
3. Any shared content should be made available. In particular, the title of any presentations should be announced and the slides available to download for the user to read.

G5. The provision of audio assistance for various tasks.

Tasks such as determining whether your face is clearly visible and appropriately framed by the video camera are difficult for screen reader users and will therefore require technological assistance. This will generally involve audio input. Recommendations include:

1. The availability of support to properly frame the user’s face in the camera, including audible cues about its position relative to the screen or edge e.g. “your face is near the top edge” and feedback about background images and camera clarity. Face recognition software will be required. However, this could raise other issues, discussion of which is beyond the scope of this paper, related to the potential for recording the images of blind or other participants.
2. The use of optical character recognition (OCR) to provide information on screen content shared by speakers.
3. Mechanisms for managing (simultaneous) audio output from speakers and the screen reader. Options include left and right headphone channels or speakers for the two inputs and a function or screen reader command to switch between the two audios and regulate the volume of the SR audio independently of the main meeting.

4. A search function for the chat and participant lists to enable participants to search for e.g. chat messages from particular individuals and particular topics.

G6. Options for personalising the user interface and function and commands operation

Personalisation options allow the user to set up the system to better meet their needs, for instance by determining what notifications are read automatically and which need to be turned on.

1. The option to turn notifications on and off and customize their format avoids irritation and allows users to choose whether to focus on the meeting or screen reader messages. This should include the option to turn on/off short sounds to indicate e.g. chat messages, raised hands and participants entering and leaving, and the additional information provided and its format e.g. names and text of chat messages.
2. Users should be able to customize shortcuts and gestures to avoid possible conflicts with other programs and screen reader commands and make them easy to remember and use.

8.3 Examples

In the following two examples are introduced to illustrate how improvements can be applied.

Chat

Access to the chat can be improved in various ways, including the following:

- a. A chat block or panel, which can be accessed through both keyboard shortcuts and standard focus handling via keyboard with Tab, Ctrl+Tab or F6 keys.
- b. Arranging the messages in a list that the focus can be directed to and can be navigated with the up and down arrow keys.
- c. Each message should comprise the writer's name followed by the text with the name only read the first time for multiple consecutive messages from the same participant.
- d. Additional information, such as the time and delivery/read status should be displayed after the message text to facilitate skipping without missing the message text and the option to turn on and off or skip over this information.
- e. The option to search the list of messages for e.g. messages from a particular person or on a particular topic.
- f. The option to turn on and off automatic reading of incoming chat messages. This can be very useful and avoid having to move the focus when the focus is elsewhere, but could make it difficult to focus on speakers, particularly if their contributions involve technical or other details.
- g. Using a short and distinct sound rather than an audio message such as "message sent" to indicate a message has been sent.
- h. The option to use keyboard shortcuts or gestures to copy messages and open links for making this easier to do.

Participant information

Easy access to information about other participants e.g. number, names, camera and microphone status should include:

- a. A participant information block or panel that can be easily accessed via a keyboard shortcut or standard keyboard focus handling with Tab, Ctrl+Tab or F6 key.

- b. A list of participants which the focus can be directed to and can be navigated with the up and down arrow keys.
- c. Including the 'name', followed by camera and microphone on/off status e.g. Alice Smith, mic on, cam off" and "Bob, hand raised, mic on, cam off".
- d. Putting participants with raised hands at the top of the list to make it easier for screen reader users chairing a meeting to find and read out their names and invite them to speak.
- e. An option to search the list for e.g. a particular participant or all participants with cameras on.
- f. A filter to choose to display only e.g. participants with raised hands or camera or microphone on.

9 CONCLUSIONS

This study has investigated the accessibility and usability of three popular video conferencing tools, Zoom, Google Meet and MS Teams, for blind people interacting via screen reader. It includes an inspection evaluation of the nine functions required for the participant role and two online surveys of tool use on desktop and mobile devices, which received 65 and 94 responses respectively.

To conclude, we are able to briefly answer our two research questions:

- 1) To what extent are screen reader users able to use all the features of video conferencing tools or do they experience barriers to accessing some important functions via screen reader?
- 2) How easy and satisfying is it for screen reader users to use video conferencing tools? In particular, is interaction easy or do users experience cognitive overload and a poor and time-consuming interaction?

As stated in the discussion and shown in table 7, all three tools provided means of accessing the main, but not all the functions. Even Zoom, which overall performed the best, was not fully accessible. Zoom and Meet had better accessibility on desktop computers, whereas that for Teams was better on mobile devices. In addition, participant comments indicated that some participants were unable to use functions, such as accessing and using the chat and checking whether other participants had a raised hand, which were accessible via keyboard and screen reader according to the inspection evaluation.

The results and in particular tables 7 and 8 showed considerable variation in satisfaction for the three tools, with satisfaction highest on Zoom on both desktop (81.5%) and mobile (73.8%) devices followed by 57.6% satisfied or very satisfied with Teams on desktop and 37.5% with Meet on mobile devices. This shows that significant percentages of participants were dissatisfied with Teams and Meet and that satisfaction with Zoom was not universal. Table 5 and 6 show that none of the tools was particularly easy to use on either desktops (without keyboard shortcuts) or mobile devices. Zoom was easiest to use on desktops, but only just under a third of participants found it easy or very easy to use and Teams on mobile devices (57%). The majority of participants found Zoom and Meet neither easy nor difficult to use on both desktops and mobile devices. Participant comments show that the process was often difficult and time consuming, particularly where shortcut keys were not available. The recommendations in section 8.1 indicate some of the changes required to make tool use easier, more satisfying and less stressful and cognitively demanding for users accessing them via keyboard and screenreader.

The survey results confirmed the difficulties detected by the inspection evaluation. The results indicated that Zoom was the most used videoconference tool on both desktop and mobile devices. Teams was the least used tool and mainly adopted by technologically skilled people. Participants commented that it is very difficult to use and some users needed the help of a sighted person to understand the UI (though they were expert users). However, after getting over their initial difficulties, they considered Teams to be powerful and relatively easy to use. This agrees with the results of previous studies. The level of SR complexity makes it difficult for blind users to develop adequate mental models without

assistance from sighted people [35] and [36]. Mental models are essential in supporting users to use commands as well as interpret system actions and feedback [45].

Computers were the preferred device for accessing all the video conferencing tools, probably due to the availability of keyboard interaction and to the ability to switch easily between other applications (for taking notes, reading email, accessing instant messages) that would be complicated on smartphones/tablets.

The combination of expert evaluation and end user surveys had a number of advantages. However, the study also had several limitations. In particular, the expert evaluation was carried out by one blind and one sighted expert rather than two blind experts, few of the participants had used Teams, and the results only cover the participant and not also the host role. In addition, we did not consider the different applications and contexts in which participants were using the tools, including the types of applications, the particular screen reader used and the operating system, as well as fact that we did not include an answer option for participants to indicate that they were not aware of a particular function on a particular tool. Most of these limitations arose from the need to keep the questionnaires a manageable length and could be investigated in further work. For instance, it might be particularly interesting to investigate disabled people's experiences as meeting hosts and the associated accessibility and usability issues.

The study highlighted both the accessibility features of the three tools and their accessibility and usability problems for screen reader users. We proposed a set of guidelines for developers of both video conferencing tools and screen reader assistive technologies. Screen reader users need to be able to access all relevant information about the interface components and events and the screen reader needs to be able to obtain appropriate information from the UI, and provide appropriate feedback to the user. The guidelines are aimed at bridging the gap between current tools and the requirements of screen reader users in order to improve their experiences and make it easier for them to use video conferencing tools.

The results of the study and guidelines should give developers a better understanding of what to consider when designing applications and lead to better screen reader user experiences. Further work will focus on the evaluation methodology applied in this study which involved both blind and sighted accessibility experts in the inspection evaluation. This will lead to a better understanding of whether and, if so, how it can most effectively be applied to help designers and others in testing tool accessibility. Though beyond the remit of this paper, there is also a need for investigation of how more blind people can be encouraged to become accessibility experts and the training and other support required.

REFERENCES

- [1] Abascal, J., Arrue, M., & Valencia, X. (2019). Tools for web accessibility evaluation. In *Web Accessibility* (pp. 479-503). Springer, London.
- [2] Acosta-Vargas, P., Guaña-Moya, J., Acosta-Vargas, G., Villegas-Ch, W., & Salvador-Ullauri, L. (2021, February). Method for Assessing Accessibility in Videoconference Systems. In *International Conference on Intelligent Human Systems Integration* (pp. 669-675). Springer, Cham.
- [3] Alsaeedi, A. (2020). Comparing web accessibility evaluation tools and evaluating the accessibility of webpages: proposed frameworks. *Information*, 11(1), 40.
- [4] Anderson, N. (2021, July). Accessibility Challenges of Video Conferencing Technology. In *International Conference on Human-Computer Interaction* (pp. 185-194). Springer, Cham.
- [5] Apple Support (2022). Learn VoiceOver gestures on iPhone, <https://support.apple.com>. Accessed on January 28, 2022
- [6] Apple support (2022). VoiceOver Commands and Gestures for MAC, https://www.apple.com/voiceover/info/guide/_1131.html. Accessed on January 28, 2022
- [7] Barada, V., Doolan, K., Burić, I., Krolo, K., & Tonković, Ž. (2020). Student life during the COVID-19 pandemic lockdown: Europe-Wide Insights. *University of Zadar*.
- [8] Barbosa, T. J., & Barbosa, M. J. (2019). Zoom: An Innovative Solution For The Live-Online Virtual Classroom. *HETS Online Journal*, 9(2).
- [9] Bleakley, A., Rough, D., Edwards, J., Doyle, P., Dumbleton, O., Clark, L., ... & Cowan, B. R. (2021). Bridging social distance during social distancing: exploring social talk and remote collegiality in video conferencing. *Human-Computer Interaction*, 1-29.
- [10] Bornemann-Jeske, B. (1996, July). Usability tests on computer access devices for the blind and visually impaired. In *Interdisciplinary Aspects on Computers Helping People with Special Needs, Proceedings of the 5th International Conference ICCHP* (Vol. 96, pp. 139-147).
- [11] Borodin, Y., Bigham, J. P., Dausch, G., & Ramakrishnan, I. V. (2010, April). More than meets the eye: a survey of screen-reader browsing strategies. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)* (pp. 1-10).
- [12] Brinkley, J., & Tabrizi, N. (2017, September). A desktop usability evaluation of the facebook mobile interface using the jaws screen reader with blind users. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 61, No. 1, pp. 828-832). Sage CA: Los Angeles, CA: SAGE Publications.
- [13] Burgan, P. (2021). *The Trajectory of Zoom: Analyzing the Development of Video Conferencing Software and Accessibility in an Age of Remote Work*.
- [14] Buzzi, M. C., Buzzi, M., Leporini, B., Mori, G., & Penichet, V. M. (2010, July). Accessing Google docs via screen reader. In *International Conference on Computers for Handicapped Persons* (pp. 92-99). Springer
- [15] Buzzi, M. C., Buzzi, M., Leporini, B., & Trujillo, A. (2017). Analyzing visually impaired people's touch gestures on smartphones. *Multimedia Tools and Applications*, 76(4), 5141-5169.
- [16] Calvo, R., Seyedarabi, F., & Savva, A. (2016, December). Beyond web content accessibility guidelines: Expert accessibility reviews. In *Proceedings of the 7th international conference on software development and technologies for enhancing accessibility and fighting info-exclusion* (pp. 77-84).
- [17] Camilleri, M. A., & Camilleri, A. (2022). Remote learning via video conferencing technologies: Implications for research and practice. *Technology in Society*, 101881.
- [18] Carvalho, M. C. N., Dias, F. S., Reis, A. G. S., & Freire, A. P. (2018, April). Accessibility and usability problems encountered on websites and applications in mobile devices by blind and normal-vision users. In *Proceedings of the 33rd Annual ACM symposium on applied computing* (pp. 2022-2029).
- [19] Chandrashekar, S., Stockman, T., Fels, D., & Benedyk, R. (2006, October). Using think aloud protocol with blind users: a case for inclusive usability evaluation methods. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility* (pp. 251-252).
- [20] Cicha, K., Rizun, M., Rutecka, P., & Strzelecki, A. (2021). COVID-19 and higher education: first-year students' expectations toward distance learning. *Sustainability*, 13(4), 1889.
- [21] Correia, A. P., Liu, C., & Xu, F. (2020). Evaluating videoconferencing systems for the quality of the educational experience. *Distance Education*, 41(4), 429-452.
- [22] Damaceno, R. J. P., Braga, J. C., & Mena-Chalco, J. P. (2018). Mobile device accessibility for the visually impaired: problems mapping and recommendations. *Universal Access in the Information Society*, 17(2), 421-435.
- [23] Díaz, J., Harari, I., Amadeo, A. P., Schiavoni, A., Gómez, S., & Osorio, A. (2022). Higher Education and Virtuality from an Inclusion Approach. In *Argentine Congress of Computer Science* (pp. 78-91). Springer, Cham.
- [24] Ferraz, R., & Diniz, V. (2021, June). Study on Accessibility of Videoconferencing Tools on Web Platforms. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1-6). IEEE.
- [25] Freedom Scientific (2022). JAWS Keystrokes, <https://support.freedomscientific.com/Content/Documents/Manuals/JAWS/Keystrokes.pdf>. Accessed on January 28, 2022
- [26] Gonçalves, R., Rocha, T., Martins, J., Branco, F., & Au-Yong-Oliveira, M. (2018). Evaluation of e-commerce websites accessibility and usability: an e-commerce platform analysis with the inclusion of blind users. *Universal Access in the Information Society*, 17(3), 567-583.
- [27] Google support (2022). Use TalkBack gestures, <https://support.google.com/accessibility/android/answer/6151827?hl=en>. Accessed on January

- [28] Guo, A., Chen, X. A., Qi, H., White, S., Ghosh, S., Asakawa, C., & Bigham, J. P. (2016, October). Vizlens: A robust and interactive screen reader for interfaces in the real world. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 651-664).
- [29] Hersh, M., Leporini, B., & Buzzi, M. (2020, September). Accessibility Evaluation of Video Conferencing Tools to Support Disabled People in Distance Teaching, Meetings and other Activities. In *ICCHP* (p. 133).
- [30] Ivory, M. Y., Yu, S., & Gronemyer, K. (2004, April). Search result exploration: a preliminary study of blind and sighted users' decision making and performance. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1453-1456)
- [31] Johns, H., Burrows, E. L., Rethnam, V., Kramer, S., & Bernhardt, J. (2021). "Can you hear me now?" Video conference coping strategies and experience during COVID-19 and beyond. *Work*, (Preprint), 1-10.
- [32] Karlapp, M., & Köhlmann, W. (2017). Adaptation and Evaluation of a Virtual Classroom for Blind Users. *i-com*, 16(1), 45-55.
- [33] Köhlmann, W., & Lucke, U. (2015, July). Alternative concepts for accessible virtual classrooms for blind users. In *2015 IEEE 15th International Conference on Advanced Learning Technologies* (pp. 413-417). IEEE.
- [34] Kulkarni, M. (2019). Digital accessibility: Challenges and opportunities. *IIMB Management Review*, 31(1), 91-98.
- [35] Kurniawan, S. H., Sutcliffe, A. G., & Blenkhorn, P. (2003). How Blind Users' Mental Models Affect Their Perceived Usability of an Unfamiliar Screen Reader. In *InterAct* (Vol. 3, pp. 631-638).
- [36] Landau, S. (1999). Tactile Graphics and Strategies for Non-Visual Seeing. *Thresholds* 1999; (19): 78–82. https://doi.org/10.1162/thld_a_00491
- [37] Lazar, J., Allen, A., Kleinman, J., & Malarkey, C. (2007). What frustrates screen reader users on the web: A study of 100 blind users. *International Journal of human-computer interaction*, 22(3), 247-269.
- [38] Leporini, B., Andronico, P., Buzzi, M., & Castillo, C. (2008). Evaluating a modified Google user interface via screen reader. *Universal access in the information society*, 7(3), 155-175.
- [39] Leporini, B., Buzzi, M., & Hersh, M. (2021, April). Distance meetings during the covid-19 pandemic: are video conferencing tools accessible for blind people?. In *Proceedings of the 18th International Web for All Conference* (pp. 1-10).
- [40] Leporini, B., & Paternò, F. (2002). Criteria for usability of accessible web sites. In *ERCIM Workshop on User Interfaces for All* (pp. 43-55). Springer, Berlin, Heidelberg.
- [41] Mahatody, T., Sagar, M., & Kolski, C. (2010). State of the art on the cognitive walkthrough method, its variants and evolutions. *Intl. Journal of Human-Computer Interaction*, 26(8), 741-785.
- [42] Maneesaeng, N., Panyabukkana, P., & Suchato, A. (2016). Accessible video-call application on android for the blind. *Lecture Notes on Software Engineering*, 4(2), 95.
- [43] Morquin, D., Chaloo, L., & Green, M. (2019, November). Teachers' Perceptions Regarding the Use of Google Classroom and Google Docs. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 21-30). Association for the Advancement of Computing in Education (AACE).
- [44] Nguyen, M. H., Gruber, J., Fuchs, J., Marler, W., Hunsaker, A., & Hargittai, E. (2020). Changes in Digital Communication During the COVID-19 Global Pandemic: Implications for Digital Inequality and Future Research. *Social Media+ Society*, 6(3), 2056305120948255.
- [45] Norman, D. A. (1988). *The psychology of everyday things*. Basic books
- [46] NV Access (2022). NVDA command key quick reference. <https://www.nvaccess.org/files/nvdaTracAttachments/455/keycommands%20with%20laptop%20keyboard%20layout.html>. Accessed on January 28, 2022
- [47] Paddison, C., & Englefield, P. (2004). Applying heuristics to accessibility inspections. *Interacting with computers*, 16(3), 507-521.
- [48] Park, E., Han, S., Bae, H., Kim, R., Lee, S., Lim, D., & Lim, H. (2019, December). Development of Automatic Evaluation Tool for Mobile Accessibility for Android Application. In *2019 International Conference on Systems of Collaboration Big Data, Internet of Things & Security (SysCoBioTS)* (pp. 1-6). IEEE.
- [49] Powlik, J. J., & Karshmer, A. I. (2002). When accessibility meets usability. *Universal Access in the Information Society*, 1(3), 217-222.
- [50] F9Pölzer, S., Schnelle-Walka, D., Pöll, D., Heumader, P., & Miesenberger, K. (2013). Making brainstorming meetings accessible for blind users. In *AAATE Conference*.
- [51] Power, C., Freire, A., Petrie, H., & Swallow, D. (2012). Guidelines Are Only Half of the Story: Accessibility Problems Encountered by Blind Users on the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 433--442.
- [52] Rømen, D., & Svanæs, D. (2008, October). Evaluating web site accessibility: validating the WAI guidelines through usability testing with disabled users. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges (NordCHI '08)*. ACM, New York, NY, USA, 535--538.
- [53] Rui Xia Ang, J., Liu, P., McDonnell, E., & Coppola, S. (2022, April). "In this online environment, we're limited": Exploring Inclusive Video Conferencing Design for Signers. In *CHI Conference on Human Factors in Computing Systems* (pp. 1-16).
- [54] Russ, S., & Hamidi, F. (2021, April). Online learning accessibility during the COVID-19 pandemic. In *Proceedings of the 18th International Web for All Conference* (pp. 1-7).
- [55] Sabri, S., & Prasada, B. (1985). Video conferencing systems. *Proceedings of the IEEE*, 73(4), 671-688.
- [56] Serhan, D. (2020). Transitioning from face-to-face to remote learning: Students' attitudes and perceptions of using Zoom during COVID-19

pandemic. *International Journal of Technology in Education and Science*, 4(4), 335-342.

- [57] Sharif, A., Chintalapati, S. S., Wobbrock, J. O., & Reinecke, K. (2021, October). Understanding Screen-Reader Users' Experiences with Online Data Visualizations. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1-16).
- [58] Stefano, F., Borsci, S., & Stamera, G. (2010). Web usability evaluation with screen reader users: implementation of the partial concurrent thinking aloud technique. *Cognitive processing*, 11(3), 263-272.
- [59] Suci, G., Anwar, M., & Pasat, A. (2018). Virtualized Video Conferencing for eLearning. *eLearning & Software for Education*, 2.
- [60] Theofanos, M. F., & Redish, J. (2003). Bridging the gap: between accessibility and usability. *interactions*, 10(6), 36-51.
- [61] W3C (2017). Web Content Accessibility Guidelines (WCAG) 2.1, <https://www.w3.org/TR/WCAG21/>.
- [62] W3C (2020). Web Content Accessibility Guidelines (WCAG) 2.2. <https://www.w3.org/TR/WCAG22/>
- [63] W3C (2021). Accessibility of Remote Meetings. W3C First Public Working Draft 14 October 2021, available at <https://www.w3.org/TR/remote-meetings/>

APPENDIX A: TOOL SECTIONS OF THE QUESTIONNAIRES

The questions here are presented for Zoom. There are identical sections of both questionnaires for MS Teams and Google Meet.

Tools Section for Zoom in Questionnaire for Desktop Devices

7. Which of the following functions are you able to access on Zoom? (list of answer options to indicate all that hold):

- TurnMic on/off
- Turn cam on/off
- Raise/lower hand
- Know who has raised their hand
- Access the list of participants
- Screen sharing
- Read the chat
- Write in the chat
- File sharing
- Know my mic status
- Know my cam status
- Know the mic status of other participants
- Know the cam status of other participants

8. How often do you use keyboard shortcuts to turn on Zoom functions? (single choice answer options; never, for some functions, always)

9. How easy do you find it to use Zoom without keyboard shortcuts on a scale from 1(very difficult) to 5 (very easy)?

10. How satisfied you are with Zoom on a scale from 1 (very dissatisfied) to 5 (very satisfied)?

Tools Section for Zoom in Questionnaire for Mobile Devices

7. Which of the following functions are you able to access on Zoom? (list of answer options to indicate all that hold):

- TurnMic on/off
- Turn cam on/off
- Raise/lower hand
- Know who has raised their hand
- Access the list of participants
- Screen sharing

Read the chat

Write in the chat

File sharing

Know my mic status

Know my cam status

Know the mic status of other participants

Know the cam status of other participants

8. How do you search for Zoom functions? (Single choice answer options)

With left and right (flick's gestures)

I explore the whole screen with one finger

I look for the button in a particular screen position (e.g. the microphone in the centre of the bottom of the screen)

9. How useful from 1 (totally useless) to 5 (very useful) do you find it for the screen reader to automatically read chat messages and the names of participants entering and leaving?

10. How easy from 1 (very difficult) to 5 (very easy) do you find it to use Zoom on a smartphone/tablet?

11. How satisfied from 1 (very dissatisfied) to 5 (very satisfied) are you with Zoom?

12. Which version of Zoom do you prefer? (app on computer, smartphone/tablet, Web version, I do not know)